Introduction
class="introduction"
Distant Galaxies.

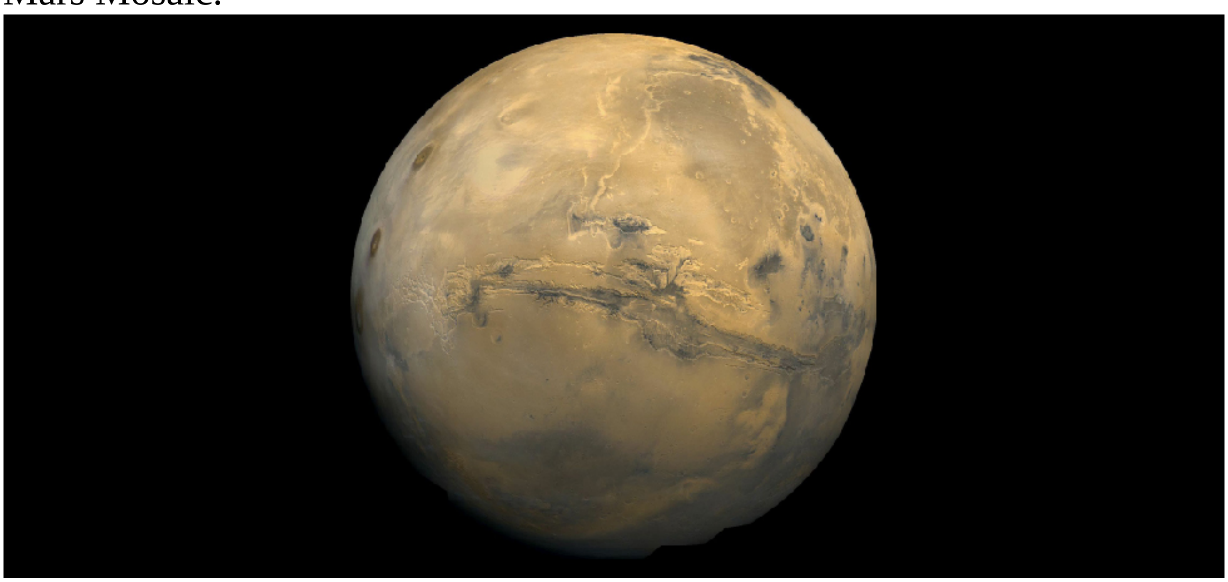
These two interacting islands of stars (galaxies) are so far away that their light takes hundreds of millions of years to reach us on Earth (photographed with the Hubble Space Telescope). (credit: modification of work by NASA, ESA, the Hubble Heritage (STScl/AURA)-ESA/Hubble Collaboration, and K. Noll (STScl))

We invite you to come along on a series of voyages to explore the universe as astronomers understand it today. Beyond Earth are vast and magnificent realms full of objects that have no counterpart on our home planet. Nevertheless, we hope to show you that the evolution of the universe has been directly responsible for your presence on Earth today.

Along your journey, you will encounter:

- a canyon system so large that, on Earth, it would stretch from Los Angeles to Washington, DC ([link]).

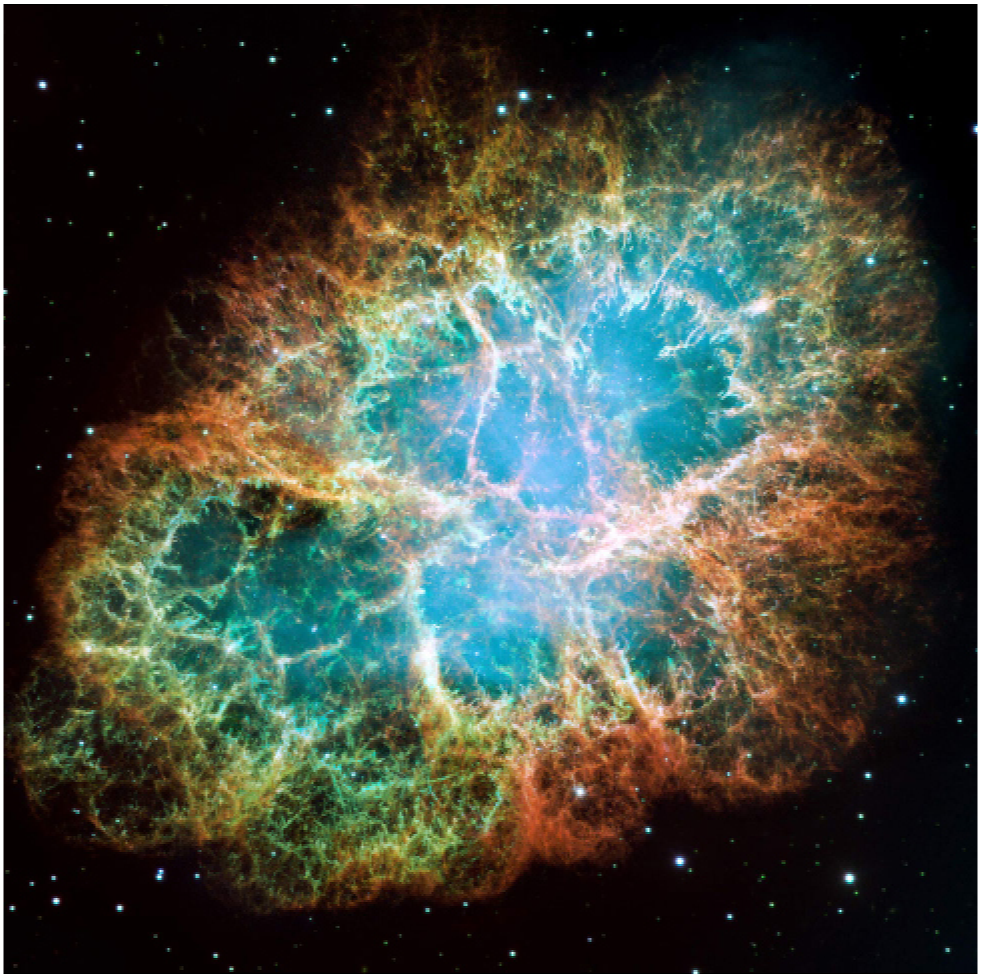
Mars Mosaic.



This image of Mars is centered on the Valles Marineris (Mariner Valley) complex of canyons, which is as long as the United States is wide. (credit: modification of work by NASA)

- a crater and other evidence on Earth that tell us that the dinosaurs (and many other creatures) died because of a cosmic collision.
- a tiny moon whose gravity is so weak that one good throw from its surface could put a baseball into orbit.
- a collapsed star so dense that to duplicate its interior we would have to squeeze every human being on Earth into a single raindrop.
- exploding stars whose violent end could wipe clean all of the life-forms on a planet orbiting a neighboring star ([link]).
- a “cannibal galaxy” that has already consumed a number of its smaller galaxy neighbors and is not yet finished finding new victims.
- a radio echo that is the faint but unmistakable signal of the creation event for our universe.

Stellar Corpse.



We observe the remains of a star that was seen to explode in our skies in 1054 (and was, briefly, bright enough to be visible during the daytime). Today, the remnant is called the Crab Nebula and its central region is seen here. Such exploding stars are crucial to the development of life in the universe. (credit: NASA, ESA, J. Hester (Arizona State University))

Such discoveries are what make astronomy such an exciting field for scientists and many others—but you will explore much more than just the objects in our universe and the latest discoveries about them. We will pay equal attention to the *process* by which we have come to understand the realms beyond Earth and the tools we use to increase that understanding.

We gather information about the cosmos from the messages the universe sends our way. Because the stars are the fundamental building blocks of the universe, decoding the message of starlight has been a central challenge and triumph of modern astronomy. By the time you have finished reading this text, you will know a bit about how to read that message and how to understand what it is telling us.

The Nature of Science

The ultimate judge in science is always what nature itself reveals based on observations, experiments, models, and testing. Science is not merely a body of knowledge, but a *method* by which we attempt to understand nature and how it behaves. This method begins with many observations over a period of time. From the trends found through observations, scientists can *model* the particular phenomena we want to understand. Such models are always approximations of nature, subject to further testing.

As a concrete astronomical example, ancient astronomers constructed a model (partly from observations and partly from philosophical beliefs) that Earth was the center of the universe and everything moved around it in circular orbits. At first, our available observations of the Sun, Moon, and planets did fit this model; however, after further observations, the model had to be updated by adding circle after circle to represent the movements of the planets around Earth at the center. As the centuries passed and improved instruments were developed for keeping track of objects in the sky, the old model (even with a huge number of circles) could no longer explain all the observed facts. As we will see in the chapter on Observing the Sky: The Birth of Astronomy, a new model, with the Sun at the center, fit the experimental evidence better. After a period of philosophical struggle, it became accepted as our view of the universe.

When they are first proposed, new models or ideas are sometimes called *hypotheses*. You may think there can be no new hypotheses in a science such as astronomy—that everything important has already been learned. Nothing could be further from the truth. Throughout this textbook you will find discussions of recent, and occasionally still controversial, hypotheses in astronomy. For example, the significance that the huge chunks of rock and ice that hit Earth have for life on Earth itself is still debated. And while the evidence is strong that vast quantities of invisible “dark energy” make up the bulk of the universe, scientists have no convincing explanation for what the dark energy actually is. Resolving these issues will require difficult observations done at the forefront of our technology, and all such hypotheses need further testing before we incorporate them fully into our standard astronomical models.

This last point is crucial: a hypothesis must be a proposed explanation that can be *tested*. The most straightforward approach to such testing in science is to perform an experiment. If the experiment is conducted properly, its results either will agree with the predictions of the hypothesis or they will contradict it. If the experimental result is truly inconsistent with the hypothesis, a scientist must discard the hypothesis and try to develop an alternative. If the experimental result agrees with predictions, this does not necessarily prove that the hypothesis is absolutely correct; perhaps later experiments will contradict crucial parts of the hypothesis. But, the more experiments that agree with the hypothesis, the more likely we are to accept the hypothesis as a useful description of nature.

One way to think about this is to consider a scientist who was born and lives on an island where only black sheep live. Day after day the scientist encounters black sheep only, so he or she hypothesizes that all sheep are black. Although every observed sheep adds confidence to the hypothesis, the scientist only has to visit the mainland and observe one white sheep to prove the hypothesis wrong.

When you read about experiments, you probably have a mental picture of a scientist in a laboratory conducting tests or taking careful measurements. This is certainly the case for a biologist or a chemist, but what can astronomers do when our laboratory is the universe? It’s impossible to put a group of stars into a test tube or to order another comet from a scientific supply company.

As a result, astronomy is sometimes called an *observational* science; we often make our tests by observing many samples of the kind of object we want to study and noting carefully how different samples vary. New instruments and technology can let us look at astronomical objects from new perspectives and in greater detail. Our hypotheses are then judged in the light of this new information, and they pass or fail in the same way we would evaluate the result of a laboratory experiment.

Much of astronomy is also a *historical* science—meaning that what we observe has already happened in the universe and we can do nothing to change it. In the same way, a geologist cannot alter what has happened to our planet, and a paleontologist cannot bring an ancient animal back to life.

While this can make astronomy challenging, it also gives us fascinating opportunities to discover the secrets of our cosmic past.

You might compare an astronomer to a detective trying to solve a crime that occurred before the detective arrived at the scene. There is lots of evidence, but both the detective and the scientist must sift through and organize the evidence to test various hypotheses about what actually happened. And there is another way in which the scientist is like a detective: they both must prove their case. The detective must convince the district attorney, the judge, and perhaps ultimately the jury that his hypothesis is correct. Similarly, the scientist must convince colleagues, editors of journals, and ultimately a broad cross-section of other scientists that her hypothesis is provisionally correct. In both cases, one can only ask for evidence “beyond a reasonable doubt.” And sometimes new evidence will force both the detective and the scientist to revise their last hypothesis.

This self-correcting aspect of science sets it off from most human activities. Scientists spend a great deal of time questioning and challenging one another, which is why applications for project funding—as well as reports for publication in academic journals—go through an extensive process of *peer review*, which is a careful examination by other scientists in the same field. In science (after formal education and training), everyone is encouraged to improve upon experiments and to challenge any and all hypotheses. New scientists know that one of the best ways to advance their careers is to find a weakness in our current understanding of something and to correct it with a new or modified hypothesis.

This is one of the reasons science has made such dramatic progress. An undergraduate science major today knows more about science and math than did Sir Isaac Newton, one of the most renowned scientists who ever lived. Even in this introductory astronomy course, you will learn about objects and processes that no one a few generations ago even dreamed existed.

The Laws of Nature

Over centuries scientists have extracted various *scientific laws* from countless observations, hypotheses, and experiments. These scientific laws are, in a sense, the "rules" of the game that nature plays. One remarkable discovery about nature—one that underlies everything you will read about in this text—is that the same laws apply everywhere in the universe. The rules that determine the motion of stars so far away that your eye cannot see them are the same laws that determine the arc of a baseball after a batter has hit it out of the park.

Note that without the existence of such universal laws, we could not make much headway in astronomy. If each pocket of the universe had different rules, we would have little chance of interpreting what happened in other "neighborhoods." But, the consistency of the laws of nature gives us enormous power to understand distant objects without traveling to them and learning the local laws. In the same way, if every region of a country had completely different laws, it would be very difficult to carry out commerce or even to understand the behavior of people in those different regions. A consistent set of laws, though, allows us to apply what we learn or practice in one state to any other state.

This is not to say that our current scientific models and laws cannot change. New experiments and observations can lead to new, more sophisticated models—models that can include new phenomena and laws about their behavior. The general theory of relativity proposed by Albert Einstein is a perfect example of such a transformation that took place about a century ago; it led us to predict, and eventually to observe, a strange new class of objects that astronomers call *black holes*. Only the patient process of observing nature ever more carefully and precisely can demonstrate the validity of such new scientific models.

One important problem in describing scientific models has to do with the limitations of language. When we try to describe complex phenomena in everyday terms, the words themselves may not be adequate to do the job. For example, you may have heard the structure of the atom likened to a miniature solar system. While some aspects of our modern model of the

atom do remind us of planetary orbits, many other of its aspects are fundamentally different.

This problem is the reason scientists often prefer to describe their models using equations rather than words. In this book, which is designed to introduce the field of astronomy, we use mainly words to discuss what scientists have learned. We avoid complex math, but if this course piques your interest and you go on in science, more and more of your studies will involve the precise language of mathematics.

Numbers in Science

In astronomy we deal with distances on a scale you may never have thought about before, with numbers larger than any you may have encountered. We adopt two approaches that make dealing with astronomical numbers a little bit easier. First, we use a system for writing large and small numbers called *scientific notation* (or sometimes *powers-of-ten notation*). This system is very appealing because it eliminates the many zeros that can seem overwhelming to the reader. In scientific notation, if you want to write a number such as 500,000,000, you express it as 5×10^8. The small raised number after the 10, called an *exponent*, keeps track of the number of places we had to move the decimal point to the left to convert 500,000,000 to 5. If you are encountering this system for the first time or would like a refresher, we suggest you look at Appendix C and [link] for more information. The second way we try to keep numbers simple is to use a consistent set of units—the metric International System of Units, or SI (from the French *Système International d'Unités*). The metric system is summarized in Appendix D (see [link]).

Note:
Watch this brief PBS animation that explains how scientific notation works and why it's useful.

A common unit astronomers use to describe distances in the universe is a light-year, which is the distance light travels during one year. Because light always travels at the same speed, and because its speed turns out to be the fastest possible speed in the universe, it makes a good standard for keeping track of distances. You might be confused because a "light-year" seems to imply that we are measuring time, but this mix-up of time and distance is common in everyday life as well. For example, when your friend asks where the movie theater is located, you might say "about 20 minutes from downtown."

So, how many kilometers are there in a light-year? Light travels at the amazing pace of 3×10^5 kilometers per second (km/s), which makes a light-year 9.46×10^{12} kilometers. You might think that such a large unit would reach the nearest star easily, but the stars are far more remote than our imaginations might lead us to believe. Even the nearest star is 4.3 light-years away—more than 40 trillion kilometers. Other stars visible to the unaided eye are hundreds to thousands of light-years away ([link]).

Orion Nebula.

This beautiful cloud of cosmic raw material (gas and dust from which new stars and planets are being made) called the Orion Nebula is about 1400 light-years away. That’s a distance of roughly 1.34×10^{16} kilometers—a pretty big number. The gas and dust in this region are illuminated by the intense light from a few extremely energetic adolescent stars. (credit: NASA, ESA, M.

Robberto (Space Telescope Science Institute/ESA) and the Hubble Space Telescope Orion Treasury Project Team)

Example:

Scientific Notation

In 2015, the richest human being on our planet had a net worth of $79.2 billion. Some might say this is an astronomical sum of money. Express this amount in scientific notation.

Solution

$79.2 billion can be written $79,200,000,000. Expressed in scientific notation it becomes $\$7.92 \times 10^{10}$.

Example:

Getting Familiar with a Light-Year

How many kilometers are there in a light-year?

Solution

Light travels 3×10^5 km in 1 s. So, let's calculate how far it goes in a year:

- There are 60 (6×10^1) s in 1 min, and 6×10^1 min in 1 h.
- Multiply these together and you find that there are 3.6×10^3 s/h.
- Thus, light covers 3×10^5 km/s $\times$ 3.6×10^3 s/h = 1.08×10^9 km/h.
- There are 24 or 2.4×10^1 h in a day, and 365.25 (3.65×10^2) days in 1 y.
- The product of these two numbers is 8.77×10^3 h/y.
- Multiplying this by 1.08×10^9 km/h gives 9.46×10^{12} km/light-year.

That's almost 10,000,000,000,000 km that light covers in a year. To help you imagine how long this distance is, we'll mention that a string 1 light-year long could fit around the circumference of Earth 236 million times.

Consequences of Light Travel Time

There is another reason the speed of light is such a natural unit of distance for astronomers. Information about the universe comes to us almost exclusively through various forms of light, and all such light travels at the speed of light—that is, 1 light-year every year. This sets a limit on how quickly we can learn about events in the universe. If a star is 100 light-years away, the light we see from it tonight left that star 100 years ago and is just now arriving in our neighborhood. The soonest we can learn about any changes in that star is 100 years after the fact. For a star 500 light-years away, the light we detect tonight left 500 years ago and is carrying 500-year-old news.

Because many of us are accustomed to instant news from the Internet, some might find this frustrating.

“You mean, when I see that star up there,” you ask, “I won’t know what’s actually happening there for another 500 years?”

But this isn’t the most helpful way to think about the situation. For astronomers, *now* is when the light reaches us here on Earth. There is no way for us to know anything about that star (or other object) until its light reaches us.

But what at first may seem a great frustration is actually a tremendous benefit in disguise. If astronomers really want to piece together what has happened in the universe since its beginning, they must find evidence about each epoch (or period of time) of the past. Where can we find evidence today about cosmic events that occurred billions of years ago?

The delay in the arrival of light provides an answer to this question. The farther out in space we look, the longer the light has taken to get here, and the longer ago it left its place of origin. By looking billions of light-years out into space, astronomers are actually seeing billions of years into the past. In this way, we can reconstruct the history of the cosmos and get a sense of how it has evolved over time.

This is one reason why astronomers strive to build telescopes that can collect more and more of the faint light in the universe. The more light we collect, the fainter the objects we can observe. On average, fainter objects are farther away and can, therefore, tell us about periods of time even deeper in the past. Instruments such as the Hubble Space Telescope ([link]) and the Very Large Telescope in Chile (which you will learn about in the chapter on Astronomical Instruments), are giving astronomers views of deep space and deep time better than any we have had before.

Telescope in Orbit.

The Hubble Space Telescope, shown here in orbit around Earth, is one of many astronomical instruments in space. (credit: modification of work by European Space Agency)

A Tour of the Universe

We can now take a brief introductory tour of the universe as astronomers understand it today to get acquainted with the types of objects and distances you will encounter throughout the text. We begin at home with Earth, a nearly spherical planet about 13,000 kilometers in diameter ([link]). A space traveler entering our planetary system would easily distinguish Earth from the other planets in our solar system by the large amount of liquid water that covers some two thirds of its crust. If the traveler had equipment to receive radio or television signals, or came close enough to see the lights of our cities at night, she would soon find signs that this watery planet has sentient life.

Humanity's Home Base.

This image shows the Western hemisphere as viewed from space 35,400 kilometers (about 22,000 miles) above Earth. Data about the land surface from one satellite was combined with another satellite's data about the clouds to create the image. (credit: modification of work by R. Stockli, A. Nelson, F. Hasler, NASA/ GSFC/ NOAA/ USGS)

Our nearest astronomical neighbor is Earth's satellite, commonly called the *Moon*. [link] shows Earth and the Moon drawn to scale on the same

diagram. Notice how small we have to make these bodies to fit them on the page with the right scale. The Moon's distance from Earth is about 30 times Earth's diameter, or approximately 384,000 kilometers, and it takes about a month for the Moon to revolve around Earth. The Moon's diameter is 3476 kilometers, about one fourth the size of Earth.

Earth and Moon, Drawn to Scale.

This image shows Earth and the Moon shown to scale for both size and distance. (credit: modification of work by NASA)

Light (or radio waves) takes 1.3 seconds to travel between Earth and the Moon. If you've seen videos of the Apollo flights to the Moon, you may recall that there was a delay of about 3 seconds between the time Mission Control asked a question and the time the astronauts responded. This was not because the astronauts were thinking slowly, but rather because it took the radio waves almost 3 seconds to make the round trip.

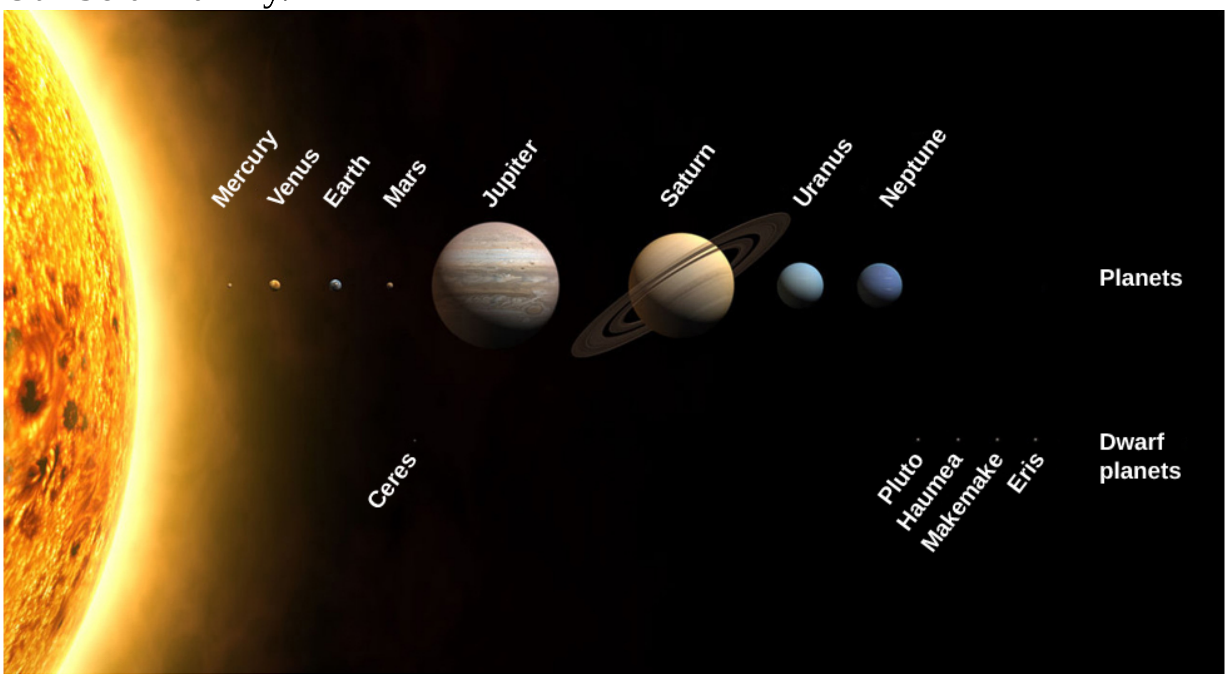
Earth revolves around our star, the Sun, which is about 150 million kilometers away—approximately 400 times as far away from us as the Moon. We call the average Earth–Sun distance an *astronomical unit* (AU) because, in the early days of astronomy, it was the most important measuring standard. Light takes slightly more than 8 minutes to travel 1 astronomical unit, which means the latest news we receive from the Sun is always 8 minutes old. The diameter of the Sun is about 1.5 million kilometers; Earth could fit comfortably inside one of the minor eruptions that occurs on the surface of our star. If the Sun were reduced to the size of a basketball, Earth would be a small apple seed about 30 meters from the ball.

It takes Earth 1 year (3×10^7 seconds) to go around the Sun at our distance; to make it around, we must travel at approximately 110,000 kilometers per hour. (If you, like many students, still prefer miles to kilometers, you might find the following trick helpful. To convert kilometers to miles, just

multiply kilometers by 0.6. Thus, 110,000 kilometers per hour becomes 66,000 miles per hour.) Because gravity holds us firmly to Earth and there is no resistance to Earth's motion in the vacuum of space, we participate in this extremely fast-moving trip without being aware of it day to day.

Earth is only one of eight planets that revolve around the Sun. These planets, along with their moons and swarms of smaller bodies such as dwarf planets, make up the solar system ([link]). A planet is defined as a body of significant size that orbits a star and does not produce its own light. (If a large body consistently produces its own light, it is then called a *star*.) Later in the book this definition will be modified a bit, but it is perfectly fine for now as you begin your voyage.

Our Solar Family.



The Sun, the planets, and some dwarf planets are shown with their sizes drawn to scale. The orbits of the planets are much more widely separated than shown in this drawing. Notice the size of Earth compared to the giant planets. (credit: modification of work by NASA)

We are able to see the nearby planets in our skies only because they reflect the light of our local star, the Sun. If the planets were much farther away,

the tiny amount of light they reflect would usually not be visible to us. The planets we have so far discovered orbiting other stars were found from the pull their gravity exerts on their parent stars, or from the light they block from their stars when they pass in front of them. We can't see most of these planets directly, although a few are now being imaged directly.

The Sun is our local star, and all the other stars are also enormous balls of glowing gas that generate vast amounts of energy by nuclear reactions deep within. We will discuss the processes that cause stars to shine in more detail later in the book. The other stars look faint only because they are so very far away. If we continue our basketball analogy, Proxima Centauri, the nearest star beyond the Sun, which is 4.3 light-years away, would be almost 7000 kilometers from the basketball.

When you look up at a star-filled sky on a clear night, all the stars visible to the unaided eye are part of a single collection of stars we call the *Milky Way Galaxy*, or simply the *Galaxy*. (When referring to the Milky Way, we capitalize *Galaxy*; when talking about other galaxies of stars, we use lowercase *galaxy*.) The Sun is one of hundreds of billions of stars that make up the Galaxy; its extent, as we will see, staggers the human imagination. Within a sphere 10 light-years in radius centered on the Sun, we find roughly ten stars. Within a sphere 100 light-years in radius, there are roughly 10,000 (10^4) stars—far too many to count or name—but we have still traversed only a tiny part of the Milky Way Galaxy. Within a 1000-light-year sphere, we find some ten million (10^7) stars; within a sphere of 100,000 light-years, we finally encompass the entire Milky Way Galaxy.

Our Galaxy looks like a giant disk with a small ball in the middle. If we could move outside our Galaxy and look down on the disk of the Milky Way from above, it would probably resemble the galaxy in [link], with its spiral structure outlined by the blue light of hot adolescent stars.

Spiral Galaxy.

This galaxy of billions of stars, called by its catalog number NGC 1073, is thought to be similar to our own Milky Way Galaxy. Here we see the giant wheel-shaped system with a bar of stars across its middle. (credit: NASA, ESA)

The Sun is somewhat less than 30,000 light-years from the center of the Galaxy, in a location with nothing much to distinguish it. From our position inside the Milky Way Galaxy, we cannot see through to its far rim (at least not with ordinary light) because the space between the stars is not completely empty. It contains a sparse distribution of gas (mostly the simplest element, hydrogen) intermixed with tiny solid particles that we call *interstellar dust*. This gas and dust collect into enormous clouds in many places in the Galaxy, becoming the raw material for future generations of stars. [link] shows an image of the disk of the Galaxy as seen from our vantage point.

Milky Way Galaxy.

Because we are inside the Milky Way Galaxy, we see its disk in cross-section flung across the sky like a great milky white avenue of stars with dark “rifts” of dust. In this dramatic image, part of it is seen above Trona Pinnacles in the California desert. (credit: Ian Norman)

Typically, the interstellar material is so extremely sparse that the space between stars is a much better vacuum than anything we can produce in

terrestrial laboratories. Yet, the dust in space, building up over thousands of light-years, can block the light of more distant stars. Like the distant buildings that disappear from our view on a smoggy day in Los Angeles, the more distant regions of the Milky Way cannot be seen behind the layers of interstellar smog. Luckily, astronomers have found that stars and raw material shine with various forms of light, some of which do penetrate the smog, and so we have been able to develop a pretty good map of the Galaxy.

Recent observations, however, have also revealed a rather surprising and disturbing fact. There appears to be more—much more—to the Galaxy than meets the eye (or the telescope). From various investigations, we have evidence that much of our Galaxy is made of material we cannot currently observe directly with our instruments. We therefore call this component of the Galaxy *dark matter*. We know the dark matter is there by the pull its gravity exerts on the stars and raw material we can observe, but what this dark matter is made of and how much of it exists remain a mystery. Furthermore, this dark matter is not confined to our Galaxy; it appears to be an important part of other star groupings as well.

By the way, not all stars live by themselves, as the Sun does. Many are born in double or triple systems with two, three, or more stars revolving about each other. Because the stars influence each other in such close systems, multiple stars allow us to measure characteristics that we cannot discern from observing single stars. In a number of places, enough stars have formed together that we recognized them as star clusters ([link]). Some of the largest of the star clusters that astronomers have cataloged contain hundreds of thousands of stars and take up volumes of space hundreds of light-years across.

Star Cluster.

This large star cluster is known by its catalog number, M9. It contains some 250,000 stars and is seen more clearly from space using the Hubble Space Telescope. It is located roughly 25,000 light-years away. (credit: NASA, ESA)

You may hear stars referred to as “eternal,” but in fact no star can last forever. Since the “business” of stars is making energy, and energy production requires some sort of fuel to be used up, eventually all stars run out of fuel. This news should not cause you to panic, though, because our Sun still has at least 5 or 6 billion years to go. Ultimately, the Sun and all stars will die, and it is in their death throes that some of the most intriguing and important processes of the universe are revealed. For example, we now

know that many of the atoms in our bodies were once inside stars. These stars exploded at the ends of their lives, recycling their material back into the reservoir of the Galaxy. In this sense, all of us are literally made of recycled “star dust.”

The Universe on the Large Scale

In a very rough sense, you could think of the solar system as your house or apartment and the Galaxy as your town, made up of many houses and buildings. In the twentieth century, astronomers were able to show that, just as our world is made up of many, many towns, so the universe is made up of enormous numbers of galaxies. (We define the universe to be everything that exists that is accessible to our observations.) Galaxies stretch as far into space as our telescopes can see, many billions of them within the reach of modern instruments. When they were first discovered, some astronomers called galaxies *island universes*, and the term is aptly descriptive; galaxies do look like islands of stars in the vast, dark seas of intergalactic space.

The nearest galaxy, discovered in 1993, is a small one that lies 75,000 light-years from the Sun in the direction of the constellation Sagittarius, where the smog in our own Galaxy makes it especially difficult to discern. (A constellation, we should note, is one of the 88 sections into which astronomers divide the sky, each named after a prominent star pattern within it.) Beyond this Sagittarius dwarf galaxy lie two other small galaxies, about 160,000 light-years away. First recorded by Magellan's crew as he sailed around the world, these are called the *Magellanic Clouds* ([link]). All three of these small galaxies are satellites of the Milky Way Galaxy, interacting with it through the force of gravity. Ultimately, all three may even be swallowed by our much larger Galaxy, as other small galaxies have been over the course of cosmic time.
Neighbor Galaxies.

This image shows both the Large Magellanic Cloud and the Small Magellanic Cloud above the telescopes of the Atacama Large Millimeter/Submillimeter Array (ALMA) in the Atacama Desert of northern Chile. (credit: ESO, C. Malin)

The nearest large galaxy is a spiral quite similar to our own, located in the constellation of Andromeda, and is thus called the Andromeda galaxy; it is also known by one of its catalog numbers, M31 ([link]). M31 is a little more than 2 million light-years away and, along with the Milky Way, is part of a small cluster of more than 50 galaxies referred to as the *Local Group*.
Closest Spiral Galaxy.

The Andromeda galaxy (M31) is a spiral-shaped collection of stars similar to our own Milky Way. (credit: Adam Evans)

At distances of 10 to 15 million light-years, we find other small galaxy groups, and then at about 50 million light-years there are more impressive systems with thousands of member galaxies. We have discovered that galaxies occur mostly in clusters, both large and small ([link]).

Fornax Cluster of Galaxies.

In this image, you can see part of a cluster of galaxies located about 60 million light-years away in the constellation of Fornax. All the objects that are not pinpoints of light in the picture are galaxies of billions of stars. (credit: ESO, J. Emerson, VISTA. Acknowledgment: Cambridge Astronomical Survey Unit)

Some of the clusters themselves form into larger groups called *superclusters*. The Local Group is part of a supercluster of galaxies, called the Virgo Supercluster, which stretches over a diameter of 110 million light-years. We are just beginning to explore the structure of the universe at these enormous scales and are already encountering some unexpected findings.

At even greater distances, where many ordinary galaxies are too dim to see, we find *quasars*. These are brilliant centers of galaxies, glowing with the light of an extraordinarily energetic process. The enormous energy of the quasars is produced by gas that is heated to a temperature of millions of degrees as it falls toward a massive black hole and swirls around it. The brilliance of quasars makes them the most distant beacons we can see in the dark oceans of space. They allow us to probe the universe 10 billion light-years away or more, and thus 10 billion years or more in the past.

With quasars we can see way back close to the Big Bang explosion that marks the beginning of time. Beyond the quasars and the most distant visible galaxies, we have detected the feeble glow of the explosion itself, filling the universe and thus coming to us from all directions in space. The discovery of this “afterglow of creation” is considered to be one of the most significant events in twentieth-century science, and we are still exploring the many things it has to tell us about the earliest times of the universe.

Measurements of the properties of galaxies and quasars in remote locations require large telescopes, sophisticated light-amplifying devices, and painstaking labor. Every clear night, at observatories around the world, astronomers and students are at work on such mysteries as the birth of new stars and the large-scale structure of the universe, fitting their results into the tapestry of our understanding.

The Universe of the Very Small

The foregoing discussion has likely impressed on you that the universe is extraordinarily large and extraordinarily empty. On average, it is 10,000 times more empty than our Galaxy. Yet, as we have seen, even the Galaxy is mostly empty space. The air we breathe has about 10^{19} atoms in each cubic centimeter—and we usually think of air as empty space. In the interstellar gas of the Galaxy, there is about one atom in every cubic centimeter. Intergalactic space is filled so sparsely that to find one atom, on average, we must search through a cubic meter of space. Most of the universe is fantastically empty; places that are dense, such as the human body, are tremendously rare.

Even our most familiar solids are mostly space. If we could take apart such a solid, piece by piece, we would eventually reach the tiny molecules from which it is formed. Molecules are the smallest particles into which any matter can be divided while still retaining its chemical properties. A molecule of water (H_2O), for example, consists of two hydrogen atoms and one oxygen atom bonded together.

Molecules, in turn, are built of atoms, which are the smallest particles of an element that can still be identified as that element. For example, an atom of gold is the smallest possible piece of gold. Nearly 100 different kinds of atoms (elements) exist in nature. Most of them are rare, and only a handful account for more than 99% of everything with which we come in contact. The most abundant elements in the cosmos today are listed in [link]; think of this table as the “greatest hits” of the universe when it comes to elements.

The Cosmically Abundant Elements

Element[footnote] This list of elements is arranged in order of the atomic number, which is the number of protons in each nucleus.	**Symbol**	**Number of Atoms per Million Hydrogen Atoms**
Hydrogen	H	1,000,000
Helium	He	80,000
Carbon	C	450
Nitrogen	N	92
Oxygen	O	740
Neon	Ne	130
Magnesium	Mg	40
Silicon	Si	37
Sulfur	S	19
Iron	Fe	32

All atoms consist of a central, positively charged nucleus surrounded by negatively charged electrons. The bulk of the matter in each atom is found in the nucleus, which consists of positive protons and electrically neutral neutrons all bound tightly together in a very small space. Each element is defined by the number of protons in its atoms. Thus, any atom with 6 protons in its nucleus is called *carbon*, any with 50 protons is called *tin*, and any with 70 protons is called *ytterbium*. (For a list of the elements, see Appendix K.)

The distance from an atomic nucleus to its electrons is typically 100,000 times the size of the nucleus itself. This is why we say that even solid matter is mostly space. The typical atom is far emptier than the solar system out to Neptune. (The distance from Earth to the Sun, for example, is only 100 times the size of the Sun.) This is one reason atoms are not like miniature solar systems.

Remarkably, physicists have discovered that everything that happens in the universe, from the smallest atomic nucleus to the largest superclusters of galaxies, can be explained through the action of only four forces: gravity, electromagnetism (which combines the actions of electricity and magnetism), and two forces that act at the nuclear level. The fact that there are four forces (and not a million, or just one) has puzzled physicists and astronomers for many years and has led to a quest for a unified picture of nature.

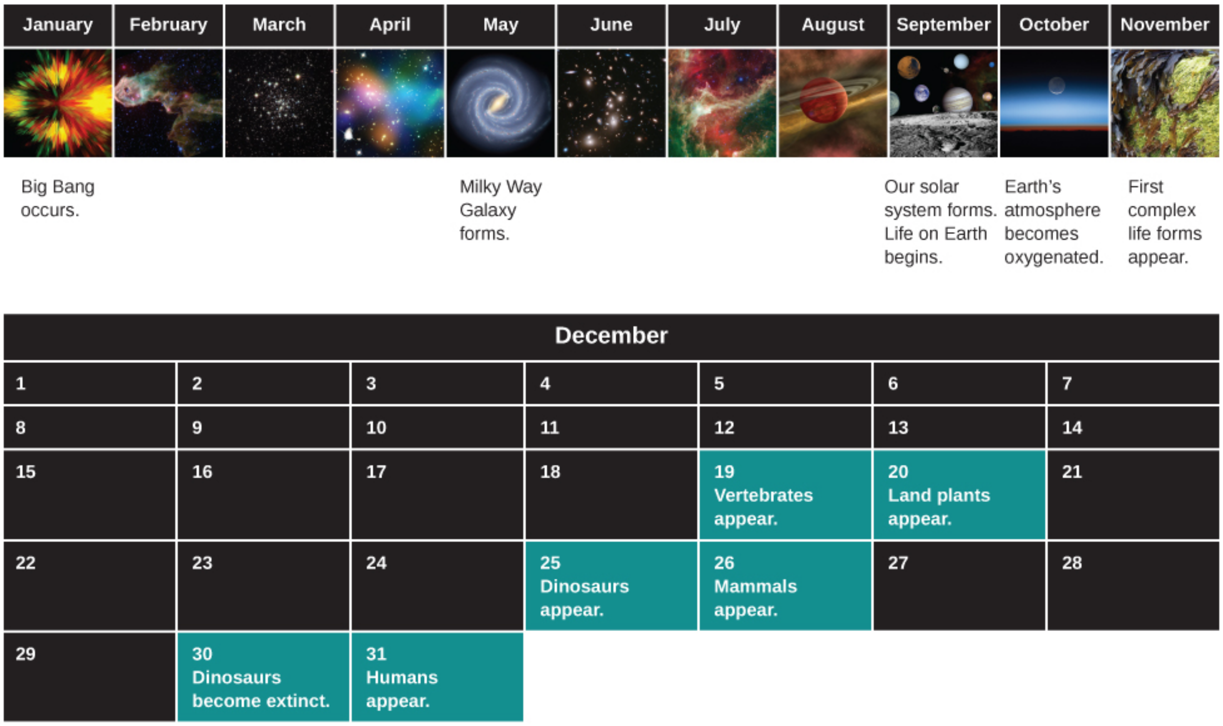
Note:
To construct an atom, particle by particle, check out this guided animation for building an atom.

A Conclusion and a Beginning

If you are new to astronomy, you have probably reached the end of our brief tour in this chapter with mixed emotions. On the one hand, you may be fascinated by some of the new ideas you’ve read about and you may be eager to learn more. On the other hand, you may be feeling a bit overwhelmed by the number of topics we have covered, and the number of new words and ideas we have introduced. Learning astronomy is a little like learning a new language: at first it seems there are so many new expressions that you’ll never master them all, but with practice, you soon develop facility with them.

At this point you may also feel a bit small and insignificant, dwarfed by the cosmic scales of distance and time. But, there is another way to look at what you have learned from our first glimpses of the cosmos. Let us consider the history of the universe from the Big Bang to today and compress it, for easy reference, into a single year. (We have borrowed this idea from Carl Sagan’s 1977 Pulitzer Prize-winning book, *The Dragons of Eden*.)

On this scale, the Big Bang happened at the first moment of January 1, and this moment, when you are reading this chapter would be the end of the very last second of December 31. When did other events in the development of the universe happen in this “cosmic year?” Our solar system formed around September 10, and the oldest rocks we can date on Earth go back to the third week in September ([link]).
Charting Cosmic Time.



On a cosmic calendar, where the time since the Big Bang is compressed into 1 year, creatures we would call human do not emerge on the scene until the evening of December 31. (credit: February: modification of work by NASA, JPL-Caltech, W. Reach (SSC/Caltech); March: modification of work by ESA, Hubble and NASA, Acknowledgement: Giles Chapdelaine; April: modification of work by NASA, ESA, CFHT, CXO, M.J. Jee (University of California, Davis), A. Mahdavi (San Francisco State University); May: modification of work by NASA, JPL-Caltech; June: modification of work by NASA/ESA; July: modification of work by NASA, JPL-Caltech, Harvard-Smithsonian; August: modification of work by NASA, JPL-Caltech, R. Hurt (SSC-Caltech); September: modification of work by NASA; October: modification of work by NASA; November: modification of work by Dénes Emőke)

Where does the origin of human beings fall during the course of this cosmic year? The answer turns out to be the evening of December 31. The invention of the alphabet doesn’t occur until the fiftieth second of 11:59

p.m. on December 31. And the beginnings of modern astronomy are a mere fraction of a second before the New Year. Seen in a cosmic context, the amount of time we have had to study the stars is minute, and our success in piecing together as much of the story as we have is remarkable.

Certainly our attempts to understand the universe are not complete. As new technologies and new ideas allow us to gather more and better data about the cosmos, our present picture of astronomy will very likely undergo many changes. Still, as you read our current progress report on the exploration of the universe, take a few minutes every once in a while just to savor how much you have already learned.

For Further Exploration

Books

Miller, Ron, and William Hartmann. *The Grand Tour: A Traveler's Guide to the Solar System*. 3rd ed. Workman, 2005. This volume for beginners is a colorfully illustrated voyage among the planets.

Sagan, Carl. *Cosmos.* Ballantine, 2013 [1980]. This tome presents a classic overview of astronomy by an astronomer who had a true gift for explaining things clearly. (You can also check out Sagan's television series *Cosmos: A Personal Voyage* and Neil DeGrasse Tyson's current series *Cosmos: A Spacetime Odyssey.*)

Tyson, Neil DeGrasse, and Don Goldsmith. *Origins: Fourteen Billion Years of Cosmic Evolution.* Norton, 2004. This book provides a guided tour through the beginnings of the universe, galaxies, stars, planets, and life.

Websites

If you enjoyed the beautiful images in this chapter (and there are many more fabulous photos to come in other chapters), you may want to know where you can obtain and download such pictures for your own enjoyment.

(Many astronomy images are from government-supported instruments or projects, paid for by tax dollars, and therefore are free of copyright laws.) Here are three resources we especially like:

- Astronomy Picture of the Day: apod.nasa.gov/apod/astropix.html. Two space scientists scour the Internet and select one beautiful astronomy image to feature each day. Their archives range widely, from images of planets and nebulae to rockets and space instruments; they also have many photos of the night sky. The search function (see the menu on the bottom of the page) works quite well for finding something specific among the many years' worth of daily images.
- Hubble Space Telescope Images: http://hubblesite.org/images/gallery. Starting at this page, you can select from among hundreds of Hubble pictures by subject or by date. Note that many of the images have supporting pictures with them, such as diagrams, animations, or comparisons. Excellent captions and background information are provided. Other ways to approach these images are through the more public-oriented Hubble Gallery (www.hubblesite.org/gallery) and the European homepage (www.spacetelescope.org/images).
- National Aeronautics and Space Administration's (NASA's) Planetary Photojournal: photojournal.jpl.nasa.gov. This site features thousands of images from planetary exploration, with captions of varied length. You can select images by world, feature name, date, or catalog number, and download images in a number of popular formats. However, only NASA mission images are included. Note the Photojournal Search option on the menu at the top of the homepage to access ways to search their archives.

Videos

Powers of Ten: www.youtube.com/watch?v=0fKBhvDjuy0. This classic short video is a much earlier version of Powers of Ten, narrated by Philip Morrison (9:00).

The Known Universe: www.youtube.com/watch?v=17jymDn0W6U. This video tour from the American Museum of Natural History has realistic

animation, music, and captions (6:30).

Wanderers: apod.nasa.gov/apod/ap141208.html. This video provides a tour of the solar system, with narrative by Carl Sagan, imagining other worlds with dramatically realistic paintings (3:50).

Thinking Ahead
class="introduction"
Night Sky.

In this panoramic photograph of the night sky from the Atacama Desert in Chile, we can see the central portion of the Milky Way Galaxy arcing upward in the center of the frame. On the left, the Large Magellanic Cloud and the Small Magellanic Cloud (smaller galaxies that orbit the Milky Way Galaxy) are easily visible from the Southern Hemisphere.

(credit: modification of work by ESO/Y. Beletsky)

Much to your surprise, a member of the Flat Earth Society moves in next door. He believes that Earth is flat and all the NASA images of a spherical Earth are either faked or simply show the round (but flat) disk of Earth from above. How could you prove to your new neighbor that Earth really is a sphere? (When you've thought about this on your own, you can check later in the chapter for some suggested answers.)

Today, few people really spend much time looking at the night sky. In ancient days, before electric lights robbed so many people of the beauty of the sky, the stars and planets were an important aspect of everyone's daily life. All the records that we have—on paper and in stone—show that ancient civilizations around the world noticed, worshipped, and tried to understand the lights in the sky and fit them into their own view of the world. These ancient observers found both majestic regularity and never-ending surprise in the motions of the heavens. Through their careful study of the planets, the Greeks and later the Romans laid the foundation of the science of astronomy.

The Sky Above

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the main features of the celestial sphere
- Explain the system astronomers use to describe the sky
- Describe how motions of the stars appear to us on Earth
- Describe how motions of the Sun, Moon, and planets appear to us on Earth
- Understand the modern meaning of the term *constellation*

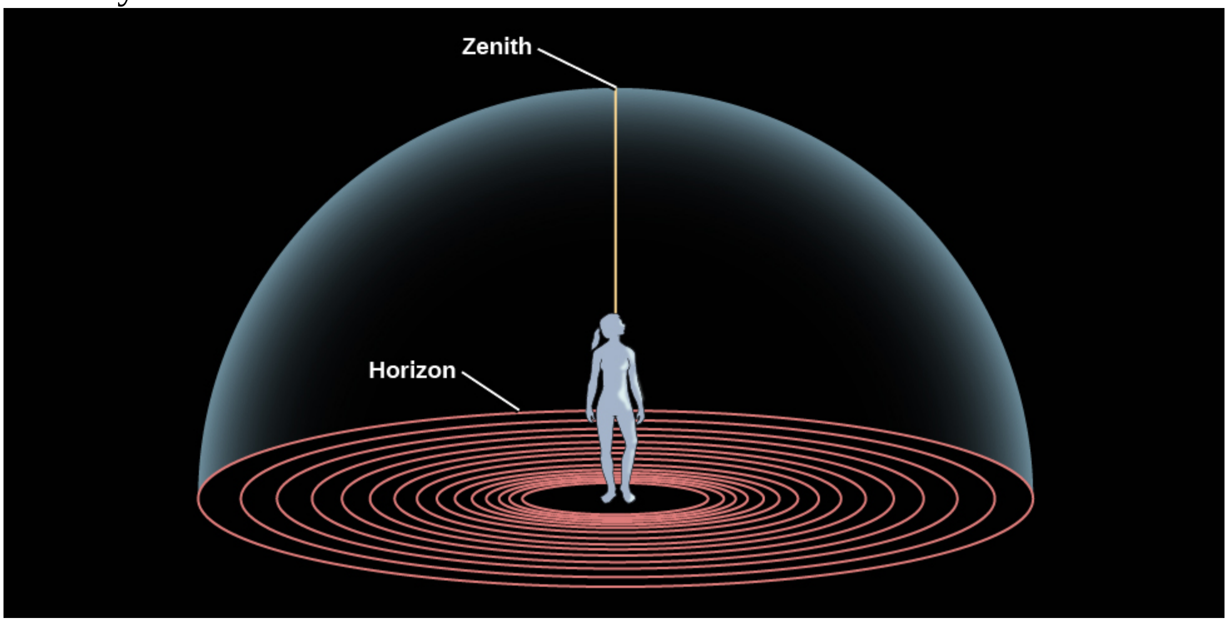
Our senses suggest to us that Earth is the center of the universe—the hub around which the heavens turn. This **geocentric** (Earth-centered) view was what almost everyone believed until the European Renaissance. After all, it is simple, logical, and seemingly self-evident. Furthermore, the geocentric perspective reinforced those philosophical and religious systems that taught the unique role of human beings as the central focus of the cosmos. However, the geocentric view happens to be wrong. One of the great themes of our intellectual history is the overthrow of the geocentric perspective. Let us, therefore, take a look at the steps by which we reevaluated the place of our world in the cosmic order.

The Celestial Sphere

If you go on a camping trip or live far from city lights, your view of the sky on a clear night is pretty much identical to that seen by people all over the world before the invention of the telescope. Gazing up, you get the

impression that the sky is a great hollow dome with you at the center ([link]), and all the stars are an equal distance from you on the surface of the dome. The top of that dome, the point directly above your head, is called the **zenith**, and where the dome meets Earth is called the **horizon**. From the sea or a flat prairie, it is easy to see the horizon as a circle around you, but from most places where people live today, the horizon is at least partially hidden by mountains, trees, buildings, or smog.

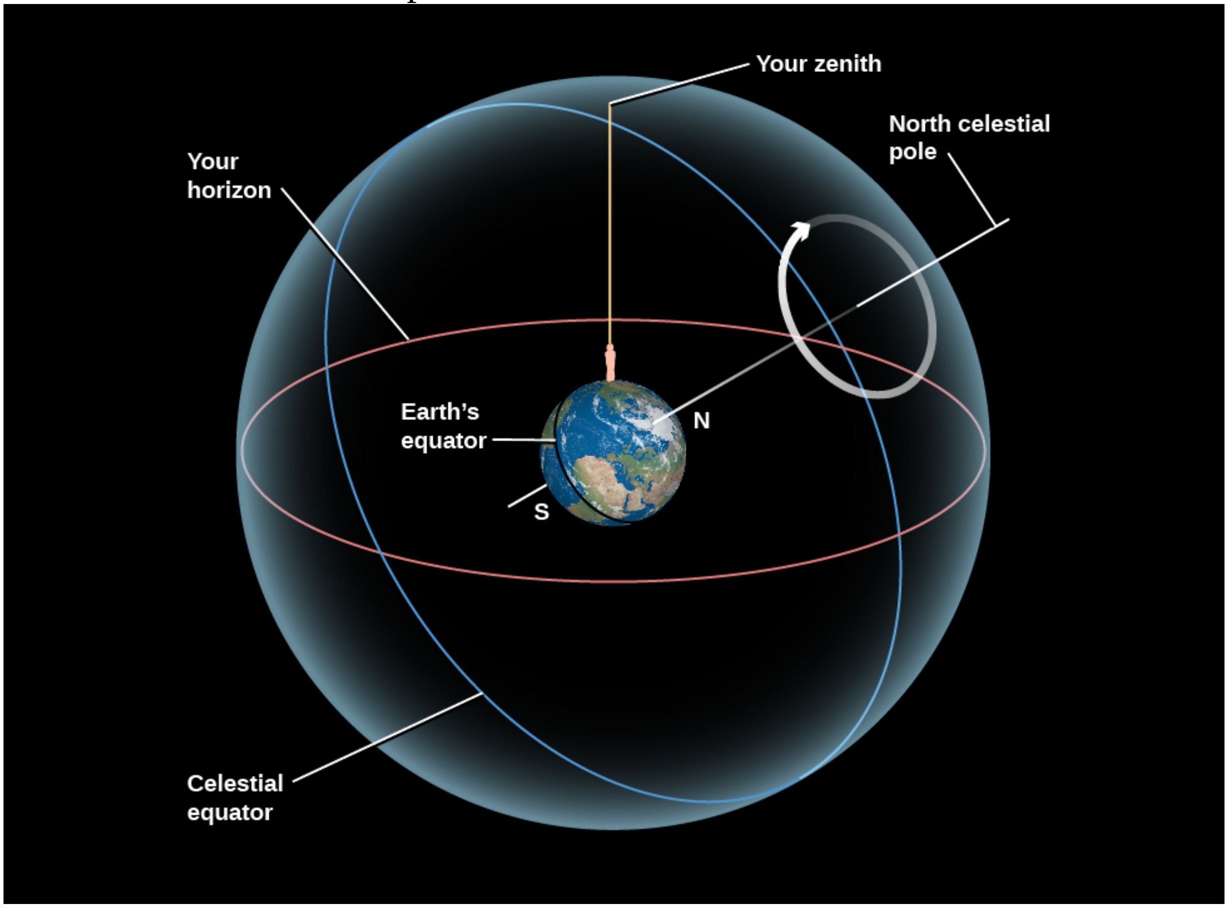
The Sky around Us.



The horizon is where the sky meets the ground; an observer's zenith is the point directly overhead.

If you lie back in an open field and observe the night sky for hours, as ancient shepherds and travelers regularly did, you will see stars rising on the eastern horizon (just as the Sun and Moon do), moving across the dome of the sky in the course of the night, and setting on the western horizon. Watching the sky turn like this night after night, you might eventually get the idea that the dome of the sky is really part of a great sphere that is turning around you, bringing different stars into view as it turns. The early Greeks regarded the sky as just such a **celestial sphere** ([link]). Some thought of it as an actual sphere of transparent crystalline material, with the stars embedded in it like tiny jewels.

Circles on the Celestial Sphere.



Here we show the (imaginary) celestial sphere around Earth, on which objects are fixed, and which rotates around Earth on an axis. In reality, it is Earth that turns around this axis, creating the illusion that the sky revolves around us. Note that Earth in this picture has been tilted so that your location is at the top and the North Pole is where the N is. The apparent motion of celestial objects in the sky around the pole is shown by the circular arrow.

Today, we know that it is not the celestial sphere that turns as night and day proceed, but rather the planet on which we live. We can put an imaginary stick through Earth's North and South Poles, representing our planet's axis. It is because Earth turns on this axis every 24 hours that we see the Sun, Moon, and stars rise and set with clockwork regularity. Today, we know

that these celestial objects are not really on a dome, but at greatly varying distances from us in space. Nevertheless, it is sometimes still convenient to talk about the celestial dome or sphere to help us keep track of objects in the sky. There is even a special theater, called a *planetarium*, in which we project a simulation of the stars and planets onto a white dome.

As the celestial sphere rotates, the objects on it maintain their positions with respect to one another. A grouping of stars such as the Big Dipper has the same shape during the course of the night, although it turns with the sky. During a single night, even objects we know to have significant motions of their own, such as the nearby planets, seem fixed relative to the stars. Only meteors—brief “shooting stars” that flash into view for just a few seconds—move appreciably with respect to other objects on the celestial sphere. (This is because they are not stars at all. Rather, they are small pieces of cosmic dust, burning up as they hit Earth’s atmosphere.) We can use the fact that the entire celestial sphere seems to turn together to help us set up systems for keeping track of what things are visible in the sky and where they happen to be at a given time.

Celestial Poles and Celestial Equator

To help orient us in the turning sky, astronomers use a system that extends Earth’s axis points into the sky. Imagine a line going through Earth, connecting the North and South Poles. This is Earth’s axis, and Earth rotates about this line. If we extend this imaginary line outward from Earth, the points where this line intersects the celestial sphere are called the *north celestial pole* and the *south celestial pole*. As Earth rotates about its axis, the sky appears to turn in the opposite direction around those **celestial poles** ([link]). We also (in our imagination) throw Earth’s equator onto the sky and call this the **celestial equator**. It lies halfway between the celestial poles, just as Earth’s equator lies halfway between our planet’s poles.
Circling the South Celestial Pole.

This long-exposure photo shows trails left by stars as a result of the apparent rotation of the celestial sphere around the south celestial pole. (In reality, it is Earth that rotates.) (Credit: ESO/Iztok Bončina)

Now let’s imagine how riding on different parts of our spinning Earth affects our view of the sky. The apparent motion of the celestial sphere depends on your latitude (position north or south of the equator). First of all, notice that Earth’s axis is pointing at the celestial poles, so these two points in the sky do not appear to turn.

If you stood at the North Pole of Earth, for example, you would see the north celestial pole overhead, at your zenith. The celestial equator, 90° from the celestial poles, would lie along your horizon. As you watched the stars during the course of the night, they would all circle around the celestial pole, with none rising or setting. Only that half of the sky north of the

celestial equator is ever visible to an observer at the North Pole. Similarly, an observer at the South Pole would see only the southern half of the sky.

If you were at Earth's equator, on the other hand, you see the celestial equator (which, after all, is just an "extension" of Earth's equator) pass overhead through your zenith. The celestial poles, being 90° from the celestial equator, must then be at the north and south points on your horizon. As the sky turns, all stars rise and set; they move straight up from the east side of the horizon and set straight down on the west side. During a 24-hour period, all stars are above the horizon exactly half the time. (Of course, during some of those hours, the Sun is too bright for us to see them.)

What would an observer in the latitudes of the United States or Europe see? Remember, we are neither at Earth's pole nor at the equator, but in between them. For those in the continental United States and Europe, the north celestial pole is neither overhead nor on the horizon, but in between. It appears above the northern horizon at an angular height, or altitude, equal to the observer's latitude. In San Francisco, for example, where the latitude is 38° N, the north celestial pole is 38° above the northern horizon.

For an observer at 38° N latitude, the south celestial pole is 38° below the southern horizon and, thus, never visible. As Earth turns, the whole sky seems to pivot about the north celestial pole. For this observer, stars within 38° of the North Pole can never set. They are always above the horizon, day and night. This part of the sky is called the north **circumpolar zone**. For observers in the continental United States, the Big Dipper, Little Dipper, and Cassiopeia are examples of star groups in the north circumpolar zone. On the other hand, stars within 38° of the south celestial pole never rise. That part of the sky is the south circumpolar zone. To most U.S. observers, the Southern Cross is in that zone. (Don't worry if you are not familiar with the star groups just mentioned; we will introduce them more formally later on.)

Note:

The Rotating Sky Lab created by the University of Nebraska–Lincoln provides an interactive demonstration that introduces the horizon coordinate system, the apparent rotation of the sky, and allows for exploration of the relationship between the horizon and celestial equatorial coordinate systems.

At this particular time in Earth's history, there happens to be a star very close to the north celestial pole. It is called Polaris, the pole star, and has the distinction of being the star that moves the least amount as the northern sky turns each day. Because it moved so little while the other stars moved much more, it played a special role in the mythology of several Native American tribes, for example (some called it the "fastener of the sky").

Note:
What's Your Angle?
Astronomers measure how far apart objects appear in the sky by using angles. By definition, there are 360° in a circle, so a circle stretching completely around the celestial sphere contains 360°. The half-sphere or dome of the sky then contains 180° from horizon to opposite horizon. Thus, if two stars are 18° apart, their separation spans about 1/10 of the dome of the sky. To give you a sense of how big a degree is, the full Moon is about half a degree across. This is about the width of your smallest finger (pinkie) seen at arm's length.

Rising and Setting of the Sun

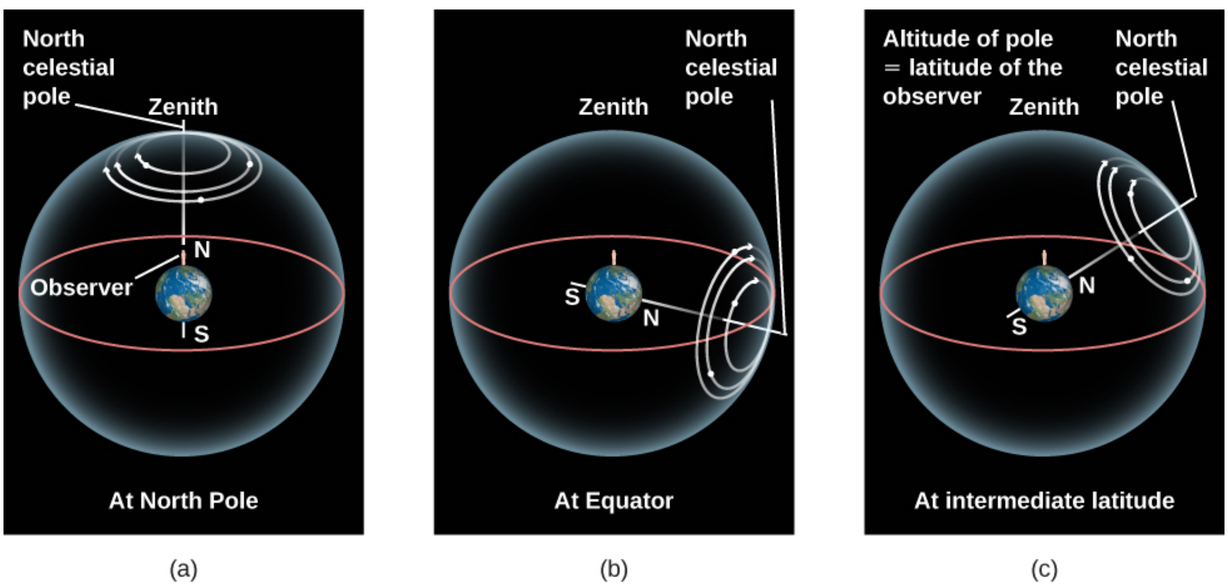
We described the movement of stars in the night sky, but what about during the daytime? The stars continue to circle during the day, but the brilliance of the Sun makes them difficult to see. (The Moon can often be seen in the daylight, however.) On any given day, we can think of the Sun as being located at some position on the hypothetical celestial sphere. When the Sun rises—that is, when the rotation of Earth carries the Sun above the horizon

—sunlight is scattered by the molecules of our atmosphere, filling our sky with light and hiding the stars above the horizon.

For thousands of years, astronomers have been aware that the Sun does more than just rise and set. It changes position gradually on the celestial sphere, moving each day about 1° to the east relative to the stars. Very reasonably, the ancients thought this meant the Sun was slowly moving around Earth, taking a period of time we call 1 **year** to make a full circle. Today, of course, we know it is Earth that is going around the Sun, but the effect is the same: the Sun’s position in our sky changes day to day. We have a similar experience when we walk around a campfire at night; we see the flames appear in front of each person seated about the fire in turn.

The path the Sun appears to take around the celestial sphere each year is called the **ecliptic** ([link]). Because of its motion on the ecliptic, the Sun rises about 4 minutes later each day with respect to the stars. Earth must make just a bit more than one complete rotation (with respect to the stars) to bring the Sun up again.

Star Circles at Different Latitudes.

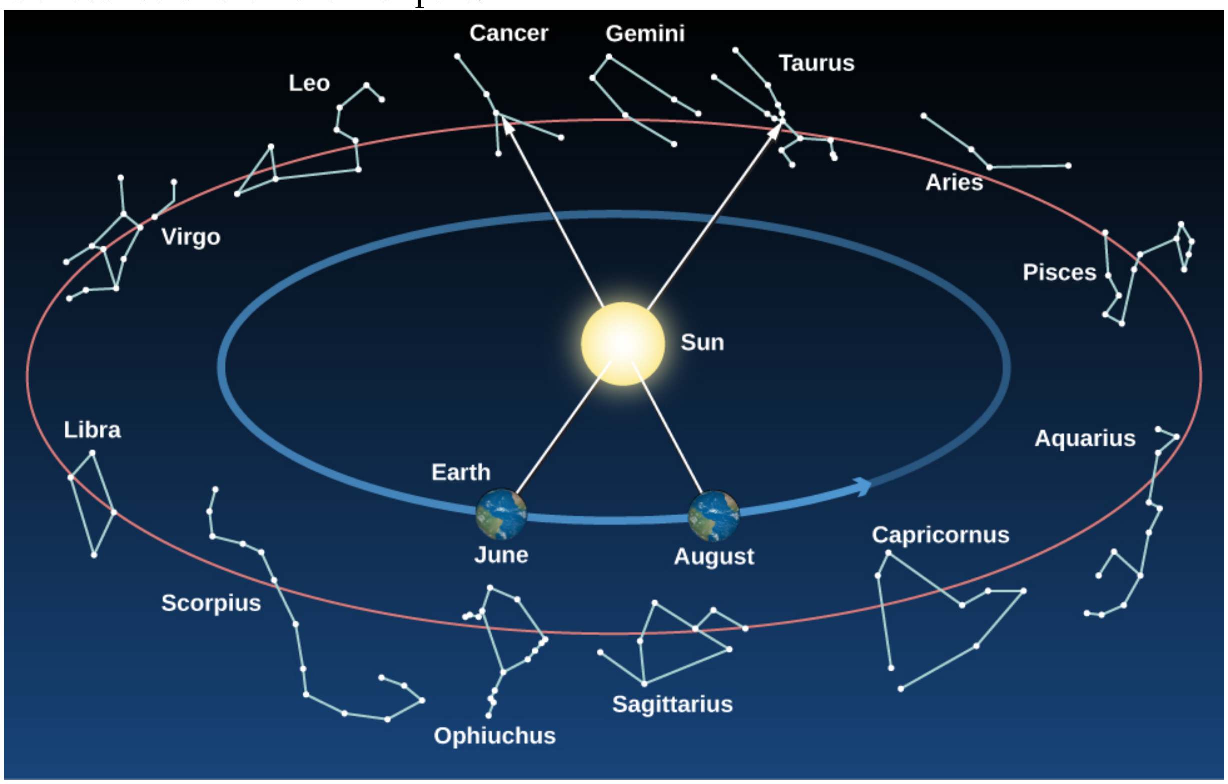


The turning of the sky looks different depending on your latitude on Earth. (a) At the North Pole, the stars circle the zenith and do not rise and set. (b) At the equator, the celestial poles are on the horizon, and the stars rise straight up and set straight down. (c) At intermediate latitudes, the north celestial pole is at some position between overhead

and the horizon. Its angle above the horizon turns out to be equal to the observer's latitude. Stars rise and set at an angle to the horizon.

As the months go by and we look at the Sun from different places in our orbit, we see it projected against different places in our orbit, and thus against different stars in the background ([link] and [link])—or we would, at least, if we could see the stars in the daytime. In practice, we must deduce which stars lie behind and beyond the Sun by observing the stars visible in the opposite direction at night. After a year, when Earth has completed one trip around the Sun, the Sun will appear to have completed one circuit of the sky along the ecliptic.

Constellations on the Ecliptic.



As Earth revolves around the Sun, we sit on "platform Earth" and see the Sun moving around the sky. The circle in the sky that the Sun appears to make around us in the course of a year is called the *ecliptic*. This circle (like all circles in the sky) goes through a set of constellations. The ancients thought these constellations, which the

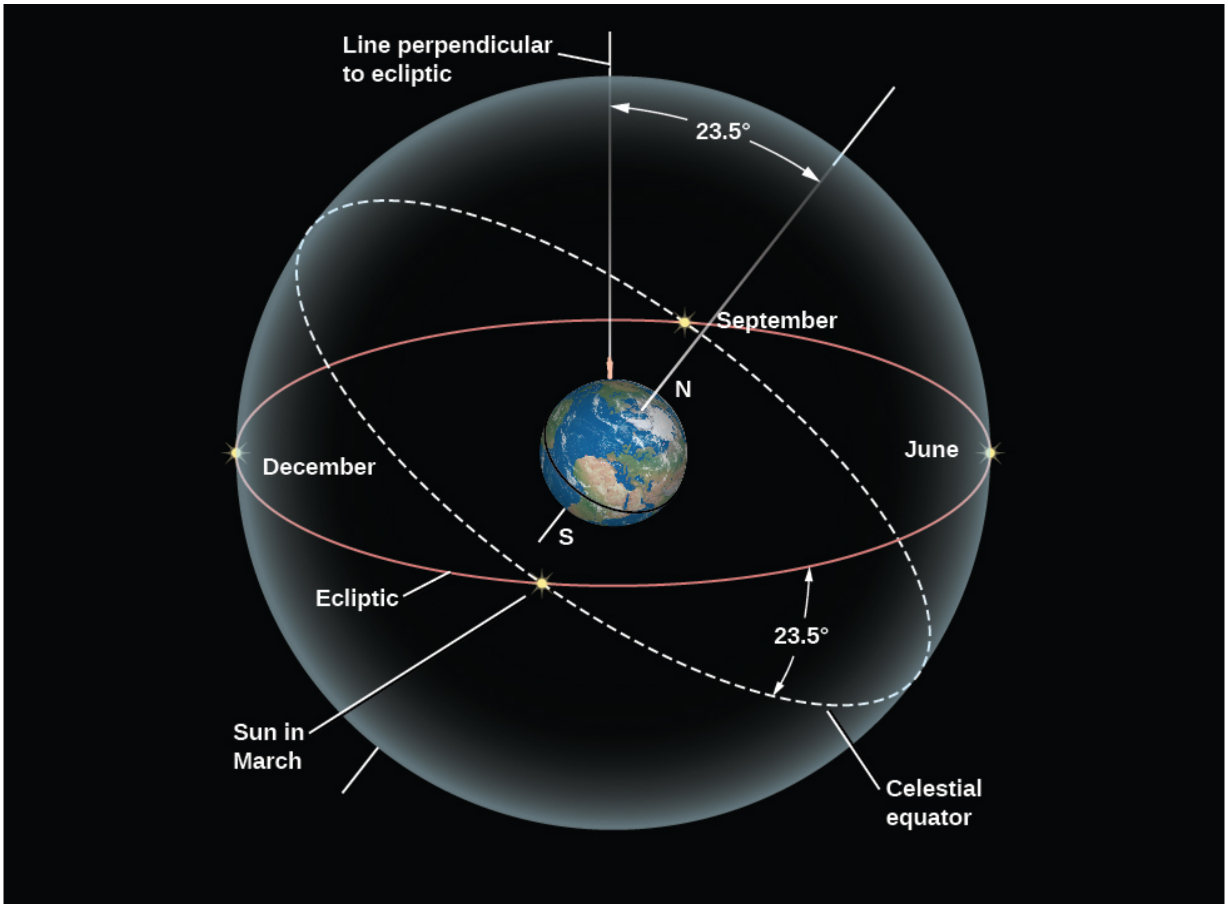
Sun (and the Moon and planets) visited, must be special and incorporated them into their system of astrology. Note that at any given time of the year, some of the constellations crossed by the ecliptic are visible in the night sky; others are in the day sky and are thus hidden by the brilliance of the Sun.

Constellations on the Ecliptic	
Constellation on the Ecliptic	**Dates When the Sun Crosses It**
Capricornus	January 21–February 16
Aquarius	February 16–March 11
Pisces	March 11–April 18
Aries	April 18–May 13
Taurus	May 13–June 22
Gemini	June 22–July 21
Cancer	July 21–August 10
Leo	August 10–September 16
Virgo	September 16–October 31
Libra	October 31–November 23

Constellations on the Ecliptic	
Constellation on the Ecliptic	**Dates When the Sun Crosses It**
Scorpius	November 23–November 29
Ophiuchus	November 29–December 18
Sagittarius	December 18–January 21

The ecliptic does not lie along the celestial equator but is inclined to it at an angle of about 23.5°. In other words, the Sun’s annual path in the sky is not linked with Earth’s equator. This is because our planet’s axis of rotation is tilted by about 23.5° from a vertical line sticking out of the plane of the ecliptic ([link]). Being tilted from “straight up” is not at all unusual among celestial bodies; Uranus and Pluto are actually tilted so much that they orbit the Sun “on their side.”

The Celestial Tilt.



The celestial equator is tilted by 23.5° to the ecliptic. As a result, North Americans and Europeans see the Sun north of the celestial equator and high in our sky in June, and south of the celestial equator and low in the sky in December.

The inclination of the ecliptic is the reason the Sun moves north and south in the sky as the seasons change. In Earth, Moon, and Sky, we discuss the progression of the seasons in more detail.

Fixed and Wandering Stars

The Sun is not the only object that moves among the fixed stars. The Moon and each of the planets that are visible to the unaided eye—Mercury, Venus, Mars, Jupiter, Saturn, and Uranus (although just barely)—also change their

positions slowly from day to day. During a single day, the Moon and planets all rise and set as Earth turns, just as the Sun and stars do. But like the Sun, they have independent motions among the stars, superimposed on the daily rotation of the celestial sphere. Noticing these motions, the Greeks of 2000 years ago distinguished between what they called the *fixed stars*—those that maintain fixed patterns among themselves through many generations—and the *wandering stars*, or **planets**. The word “planet,” in fact, means “wanderer” in ancient Greek.

Today, we do not regard the Sun and Moon as planets, but the ancients applied the term to all seven of the moving objects in the sky. Much of ancient astronomy was devoted to observing and predicting the motions of these celestial wanderers. They even dedicated a unit of time, the week, to the seven objects that move on their own; that’s why there are 7 days in a week. The Moon, being Earth’s nearest celestial neighbor, has the fastest apparent motion; it completes a trip around the sky in about 1 month (or *moonth*). To do this, the Moon moves about 12°, or 24 times its own apparent width on the sky, each day.

Example:
Angles in the Sky
A circle consists of 360 degrees (°). When we measure the angle in the sky that something moves, we can use this formula:
Equation:

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

This is true whether the motion is measured in kilometers per hour or degrees per hour; we just need to use consistent units.
As an example, let’s say you notice the bright star Sirius due south from your observing location in the Northern Hemisphere. You note the time, and then later, you note the time that Sirius sets below the horizon. You find that Sirius has traveled an angular distance of about 75° in 5 h. About how many hours will it take for Sirius to return to its original location?
Solution

The speed of Sirius is $\frac{75^{\circ}}{5\,\text{h}} = \frac{15^{\circ}}{1\,\text{h}}$. If we want to know the time required for Sirius to return to its original location, we need to wait until it goes around a full circle, or 360°. Rearranging the formula for speed we were originally given, we find:

Equation:

$$\text{time} = \frac{\text{distance}}{\text{speed}} = \frac{360^{\circ}}{15^{\circ}/\text{h}} = 24\,\text{h}$$

The actual time is a few minutes shorter than this, and we will explore why in a later chapter.

Check Your Learning

The Moon moves in the sky relative to the background stars (in addition to moving with the stars as a result of Earth’s rotation.) Go outside at night and note the position of the Moon relative to nearby stars. Repeat the observation a few hours later. How far has the Moon moved? (For reference, the diameter of the Moon is about 0.5°.) Based on your estimate of its motion, how long will it take for the Moon to return to the position relative to the stars in which you first observed it?

Note:

Answer:

The speed of the moon is 0.5°/1 h. To move a full 360°, the moon needs 720 h: $\frac{0.5^{\circ}}{1\,\text{h}} = \frac{360^{\circ}}{720\,\text{h}}$. Dividing 720 h by the conversion factor of 24 h/day reveals the lunar cycle is about 30 days.

The individual paths of the Moon and planets in the sky all lie close to the ecliptic, although not exactly on it. This is because the paths of the planets about the Sun, and of the Moon about Earth, are all in nearly the same

plane, as if they were circles on a huge sheet of paper. The planets, the Sun, and the Moon are thus always found in the sky within a narrow 18-degree-wide belt, centered on the ecliptic, called the **zodiac** ([link]). (The root of the term "zodiac" is the same as that of the word "zoo" and means a collection of animals; many of the patterns of stars within the zodiac belt reminded the ancients of animals, such as a fish or a goat.)

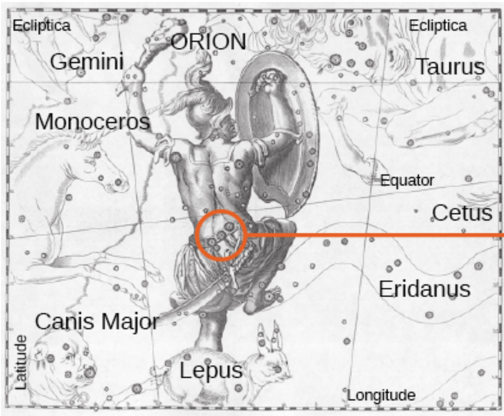
How the planets appear to move in the sky as the months pass is a combination of their actual motions plus the motion of Earth about the Sun; consequently, their paths are somewhat complex. As we will see, this complexity has fascinated and challenged astronomers for centuries.

Constellations

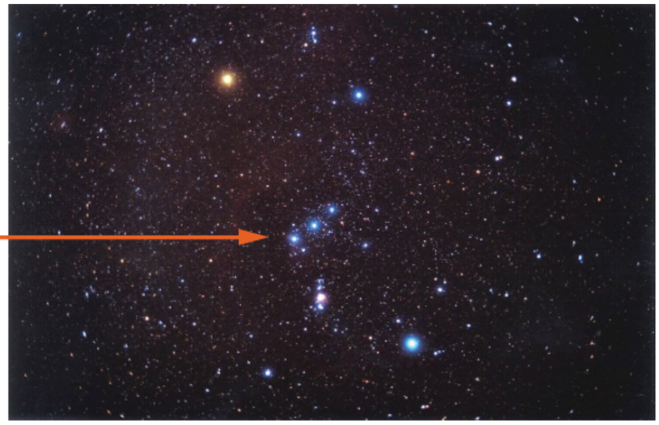
The backdrop for the motions of the "wanderers" in the sky is the canopy of stars. If there were no clouds in the sky and we were on a flat plain with nothing to obstruct our view, we could see about 3000 stars with the unaided eye. To find their way around such a multitude, the ancients found groupings of stars that made some familiar geometric pattern or (more rarely) resembled something they knew. Each civilization found its own patterns in the stars, much like a modern Rorschach test in which you are asked to discern patterns or pictures in a set of inkblots. The ancient Chinese, Egyptians, and Greeks, among others, found their own groupings—or constellations—of stars. These were helpful in navigating among the stars and in passing their star lore on to their children.

You may be familiar with some of the old star patterns we still use today, such as the Big Dipper, Little Dipper, and Orion the hunter, with his distinctive belt of three stars ([link]). However, many of the stars we see are not part of a distinctive star pattern at all, and a telescope reveals millions of stars too faint for the eye to see. Therefore, during the early decades of the 20th century, astronomers from many countries decided to establish a more formal system for organizing the sky.

Orion.



(a)


(b)

(a) The winter constellation of Orion, the hunter, is surrounded by neighboring constellations, as illustrated in the seventeenth-century atlas by Hevelius. (b) A photograph shows the Orion region in the sky. Note the three blue stars that make up the belt of the hunter. The bright red star above the belt denotes his armpit and is called Betelgeuse (pronounced “Beetel-juice”). The bright blue star below the belt is his foot and is called Rigel. (credit a: modification of work by Johannes Hevelius; b: modification of work by Matthew Spinelli)

Today, we use the term *constellation* to mean one of 88 sectors into which we divide the sky, much as the United States is divided into 50 states. The modern boundaries between the constellations are imaginary lines in the sky running north–south and east–west, so that each point in the sky falls in a specific constellation, although, like the states, not all constellations are the same size. All the constellations are listed in Appendix L. Whenever possible, we have named each modern constellation after the Latin translations of one of the ancient Greek star patterns that lies within it. Thus, the modern constellation of Orion is a kind of box on the sky, which includes, among many other objects, the stars that made up the ancient picture of the hunter. Some people use the term *asterism* to denote an especially noticeable star pattern within a constellation (or sometimes spanning parts of several constellations). For example, the Big Dipper is an asterism within the constellation of Ursa Major, the Big Bear.

Students are sometimes puzzled because the constellations seldom resemble the people or animals for which they were named. In all likelihood, the Greeks themselves did not name groupings of stars because they looked like actual people or subjects (any more than the outline of Washington state resembles George Washington). Rather, they named sections of the sky in honor of the characters in their mythology and then fit the star configurations to the animals and people as best they could.

> **Note:**
> This website about objects in the sky allows users to construct a detailed sky map showing the location and information about the Sun, Moon, planets, stars, constellations, and even satellites orbiting Earth. Begin by setting your observing location using the option in the menu in the upper right corner of the screen.

The direct evidence of our senses supports a geocentric perspective, with the celestial sphere pivoting on the celestial poles and rotating about a stationary Earth. We see only half of this sphere at one time, limited by the horizon; the point directly overhead is our zenith. The Sun's annual path on the celestial sphere is the ecliptic—a line that runs through the center of the zodiac, which is the 18-degree-wide strip of the sky within which we always find the Moon and planets. The celestial sphere is organized into 88 constellations, or sectors.

Glossary

celestial equator
: a great circle on the celestial sphere 90° from the celestial poles; where the celestial sphere intersects the plane of Earth's equator

celestial poles
: points about which the celestial sphere appears to rotate; intersections of the celestial sphere with Earth's polar axis

celestial sphere
: the apparent sphere of the sky; a sphere of large radius centered on the observer; directions of objects in the sky can be denoted by their position on the celestial sphere

circumpolar zone
: those portions of the celestial sphere near the celestial poles that are either always above or always below the horizon

ecliptic
: the apparent annual path of the Sun on the celestial sphere

geocentric
: centered on Earth

horizon (astronomical)
: a great circle on the celestial sphere 90° from the zenith; more popularly, the circle around us where the dome of the sky meets Earth

planet
: today, any of the larger objects revolving about the Sun or any similar objects that orbit other stars; in ancient times, any object that moved regularly among the fixed stars

year
: the period of revolution of Earth around the Sun

zenith
: the point on the celestial sphere opposite the direction of gravity; point directly above the observer

zodiac
: a belt around the sky about 18° wide centered on the ecliptic

Ancient Astronomy

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe early examples of astronomy around the world
- Explain how Greek astronomers were able to deduce that Earth is spherical
- Explain how Greek astronomers were able to calculate Earth's size
- Describe the motion of Earth called precession
- Describe Ptolemy's geocentric system of planetary motion

Let us now look briefly back into history. Much of modern Western civilization is derived in one way or another from the ideas of the ancient Greeks and Romans, and this is true in astronomy as well. However, many other ancient cultures also developed sophisticated systems for observing and interpreting the sky.

Astronomy around the World

Ancient Babylonian, Assyrian, and Egyptian astronomers knew the approximate length of the year. The Egyptians of 3000 years ago, for example, adopted a calendar based on a 365-day year. They kept careful track of the rising time of the bright star Sirius in the predawn sky, which has a yearly cycle that corresponded with the flooding of the Nile River. The Chinese also had a working calendar; they determined the length of the year at about the same time as the Egyptians. The Chinese also recorded

comets, bright meteors, and dark spots on the Sun. (Many types of astronomical objects were introduced in Science and the Universe: A Brief Tour. If you are not familiar with terms like *comets* and *meteors*, you may want to review that chapter.) Later, Chinese astronomers kept careful records of “guest stars”—those that are normally too faint to see but suddenly flare up to become visible to the unaided eye for a few weeks or months. We still use some of these records in studying stars that exploded a long time ago.

The Mayan culture in Mexico and Central America developed a sophisticated calendar based on the planet Venus, and they made astronomical observations from sites dedicated to this purpose a thousand years ago. The Polynesians learned to navigate by the stars over hundreds of kilometers of open ocean—a skill that enabled them to colonize new islands far away from where they began.

In Britain, before the widespread use of writing, ancient people used stones to keep track of the motions of the Sun and Moon. We still find some of the great stone circles they built for this purpose, dating from as far back as 2800 BCE. The best known of these is Stonehenge, which is discussed in Earth, Moon, and Sky.

Early Greek and Roman Cosmology

Our concept of the cosmos—its basic structure and origin—is called **cosmology**, a word with Greek roots. Before the invention of telescopes, humans had to depend on the simple evidence of their senses for a picture of the universe. The ancients developed cosmologies that combined their direct view of the heavens with a rich variety of philosophical and religious symbolism.

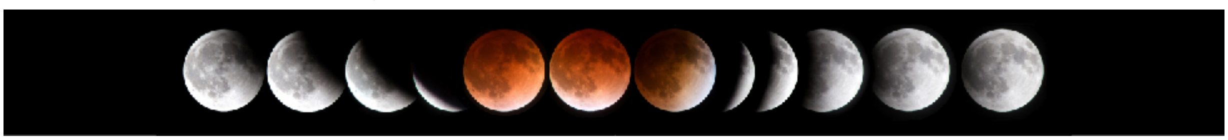
At least 2000 years before Columbus, educated people in the eastern Mediterranean region knew Earth was round. Belief in a spherical Earth may have stemmed from the time of Pythagoras, a philosopher and mathematician who lived 2500 years ago. He believed circles and spheres to be “perfect forms” and suggested that Earth should therefore be a sphere.

As evidence that the gods liked spheres, the Greeks cited the fact that the Moon is a sphere, using evidence we describe later.

The writings of Aristotle (384–322 BCE), the tutor of Alexander the Great, summarize many of the ideas of his day. They describe how the progression of the Moon's phases—its apparent changing shape—results from our seeing different portions of the Moon's sunlit hemisphere as the month goes by (see Earth, Moon, and Sky). Aristotle also knew that the Sun has to be farther away from Earth than is the Moon because occasionally the Moon passed exactly between Earth and the Sun and hid the Sun temporarily from view. We call this a *solar eclipse*.

Aristotle cited convincing arguments that Earth must be round. First is the fact that as the Moon enters or emerges from Earth's shadow during an eclipse of the Moon, the shape of the shadow seen on the Moon is always round ([link]). Only a spherical object always produces a round shadow. If Earth were a disk, for example, there would be some occasions when the sunlight would strike it edge-on and its shadow on the Moon would be a line.

Earth's Round Shadow.



A lunar eclipse occurs when the Moon moves into and out of Earth's shadow. Note the curved shape of the shadow—evidence for a spherical Earth that has been recognized since antiquity. (credit: modification of work by Brian Paczkowski)

As a second argument, Aristotle explained that travelers who go south a significant distance are able to observe stars that are not visible farther north. And the height of the North Star—the star nearest the north celestial pole—decreases as a traveler moves south. On a flat Earth, everyone would see the same stars overhead. The only possible explanation is that the traveler must have moved over a curved surface on Earth, showing stars

from a different angle. (See the How Do We Know Earth Is Round? feature for more ideas on proving Earth is round.)

One Greek thinker, Aristarchus of Samos (310–230 BCE), even suggested that Earth was moving around the Sun, but Aristotle and most of the ancient Greek scholars rejected this idea. One of the reasons for their conclusion was the thought that if Earth moved about the Sun, they would be observing the stars from different places along Earth's orbit. As Earth moved along, nearby stars should shift their positions in the sky relative to more distant stars. In a similar way, we see foreground objects appear to move against a more distant background whenever we are in motion. When we ride on a train, the trees in the foreground appear to shift their position relative to distant hills as the train rolls by. Unconsciously, we use this phenomenon all of the time to estimate distances around us.

The apparent shift in the direction of an object as a result of the motion of the observer is called **parallax**. We call the shift in the apparent direction of a star due to Earth's orbital motion *stellar parallax*. The Greeks made dedicated efforts to observe stellar parallax, even enlisting the aid of Greek soldiers with the clearest vision, but to no avail. The brighter (and presumably nearer) stars just did not seem to shift as the Greeks observed them in the spring and then again in the fall (when Earth is on the opposite side of the Sun).

This meant either that Earth was not moving or that the stars had to be so tremendously far away that the parallax shift was immeasurably small. A cosmos of such enormous extent required a leap of imagination that most ancient philosophers were not prepared to make, so they retreated to the safety of the Earth-centered view, which would dominate Western thinking for nearly two millennia.

Note:
How Do We Know Earth Is Round?
In addition to the two ways (from Aristotle's writings) discussed in this chapter, you might also reason as follows:

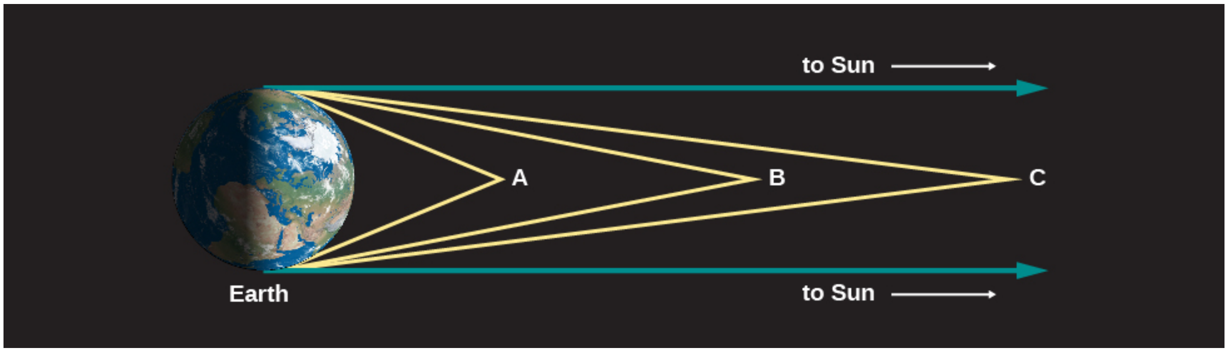
1. Let’s watch a ship leave its port and sail into the distance on a clear day. On a flat Earth, we would just see the ship get smaller and smaller as it sails away. But this isn’t what we actually observe. Instead, ships sink below the horizon, with the hull disappearing first and the mast remaining visible for a while longer. Eventually, only the top of the mast can be seen as the ship sails around the curvature of Earth. Finally, the ship disappears under the horizon.
2. The International Space Station circles Earth once every 90 minutes or so. Photographs taken from the shuttle and other satellites show that Earth is round from every perspective.
3. Suppose you made a friend in each time zone of Earth. You call all of them at the same hour and ask, “Where is the Sun?” On a flat Earth, each caller would give you roughly the same answer. But on a round Earth you would find that, for some friends, the Sun would be high in the sky whereas for others it would be rising, setting, or completely out of sight (and this last group of friends would be upset with you for waking them up).

Measurement of Earth by Eratosthenes

The Greeks not only knew Earth was round, but also they were able to measure its size. The first fairly accurate determination of Earth’s diameter was made in about 200 BCE by Eratosthenes (276–194 BCE), a Greek living in Alexandria, Egypt. His method was a geometric one, based on observations of the Sun.

The Sun is so distant from us that all the light rays that strike our planet approach us along essentially parallel lines. To see why, look at [link]. Take a source of light near Earth—say, at position A. Its rays strike different parts of Earth along diverging paths. From a light source at B, or at C (which is still farther away), the angle between rays that strike opposite parts of Earth is smaller. The more distant the source, the smaller the angle between the rays. For a source infinitely distant, the rays travel along parallel lines.

Light Rays from Space.



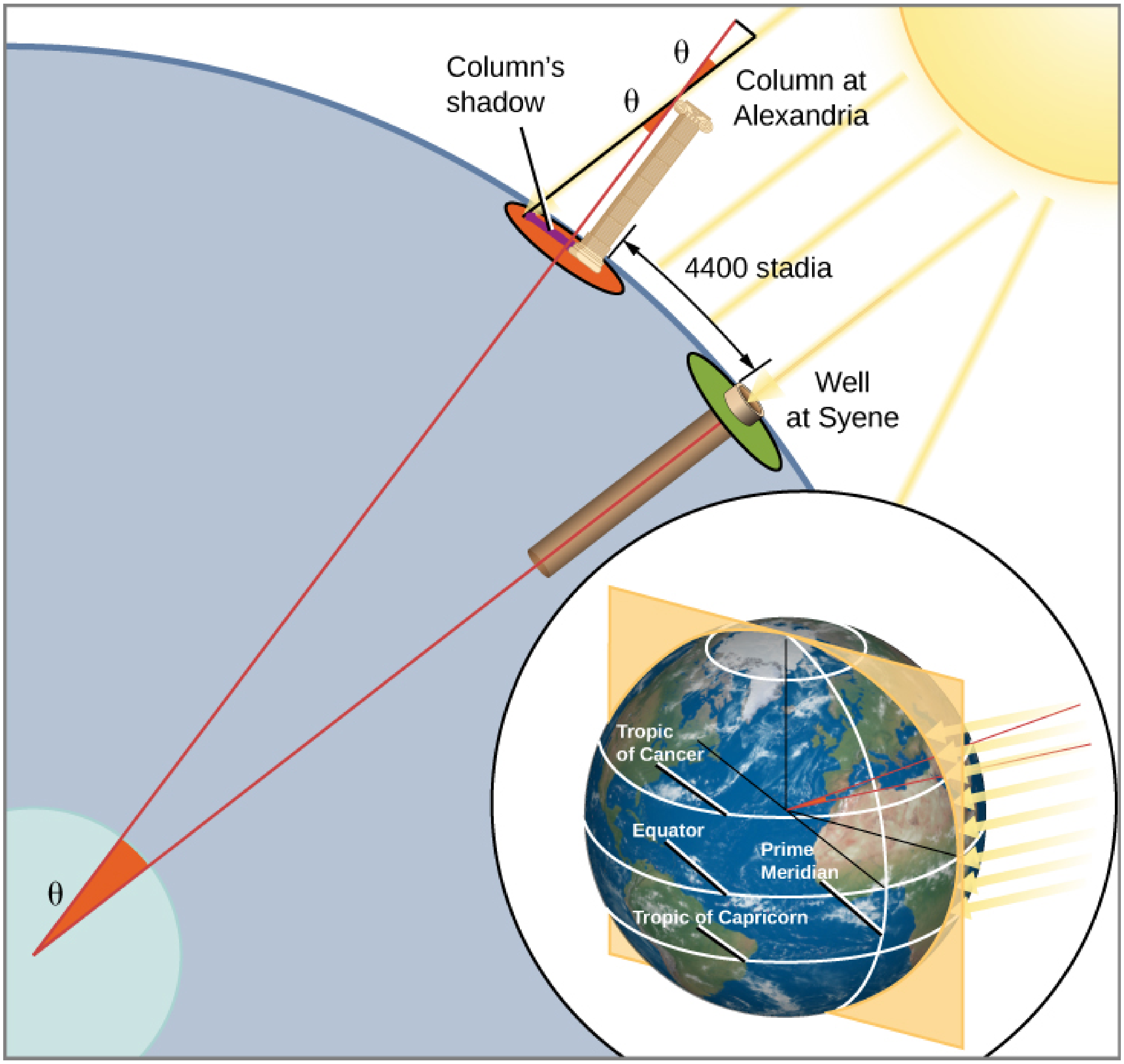
The more distant an object, the more nearly parallel the rays of light coming from it.

Of course, the Sun is not infinitely far away, but given its distance of 150 million kilometers, light rays striking Earth from a point on the Sun diverge from one another by an angle far too small to be observed with the unaided eye. As a consequence, if people all over Earth who could see the Sun were to point at it, their fingers would, essentially, all be parallel to one another. (The same is also true for the planets and stars—an idea we will use in our discussion of how telescopes work.)

Eratosthenes was told that on the first day of summer at Syene, Egypt (near modern Aswan), sunlight struck the bottom of a vertical well at noon. This indicated that the Sun was directly over the well—meaning that Syene was on a direct line from the center of Earth to the Sun. At the corresponding time and date in Alexandria, Eratosthenes observed the shadow a column made and saw that the Sun was not directly overhead, but was slightly south of the zenith, so that its rays made an angle with the vertical equal to about 1/50 of a circle (7°). Because the Sun's rays striking the two cities are parallel to one another, why would the two rays not make the same angle with Earth's surface? Eratosthenes reasoned that the curvature of the round Earth meant that "straight up" was not the same in the two cities. And the measurement of the angle in Alexandria, he realized, allowed him to figure out the size of Earth. Alexandria, he saw, must be 1/50 of Earth's circumference north of Syene ([link]). Alexandria had been measured to be 5000 stadia north of Syene. (The *stadium* was a Greek unit of length,

derived from the length of the racetrack in a stadium.) Eratosthenes thus found that Earth’s circumference must be 50 × 5000, or 250,000 stadia.

How Eratosthenes Measured the Size of Earth.



Eratosthenes measured the size of Earth by observing the angle at which the Sun’s rays hit our planet’s surface. The Sun’s rays come in parallel, but because Earth’s surface curves, a ray at Syene comes straight down whereas a ray at Alexandria makes an angle of 7° with the vertical. That means, in effect, that at Alexandria, Earth’s surface has curved away from Syene by 7° of 360°, or 1/50 of a full circle. Thus, the distance between the two cities must be 1/50 the

circumference of Earth. (credit: modification of work by NOAA Ocean Service Education)

It is not possible to evaluate precisely the accuracy of Eratosthenes solution because there is doubt about which of the various kinds of Greek stadia he used as his unit of distance. If it was the common Olympic stadium, his result is about 20% too large. According to another interpretation, he used a stadium equal to about 1/6 kilometer, in which case his figure was within 1% of the correct value of 40,000 kilometers. Even if his measurement was not exact, his success at measuring the size of our planet by using only shadows, sunlight, and the power of human thought was one of the greatest intellectual achievements in history.

Hipparchus and Precession

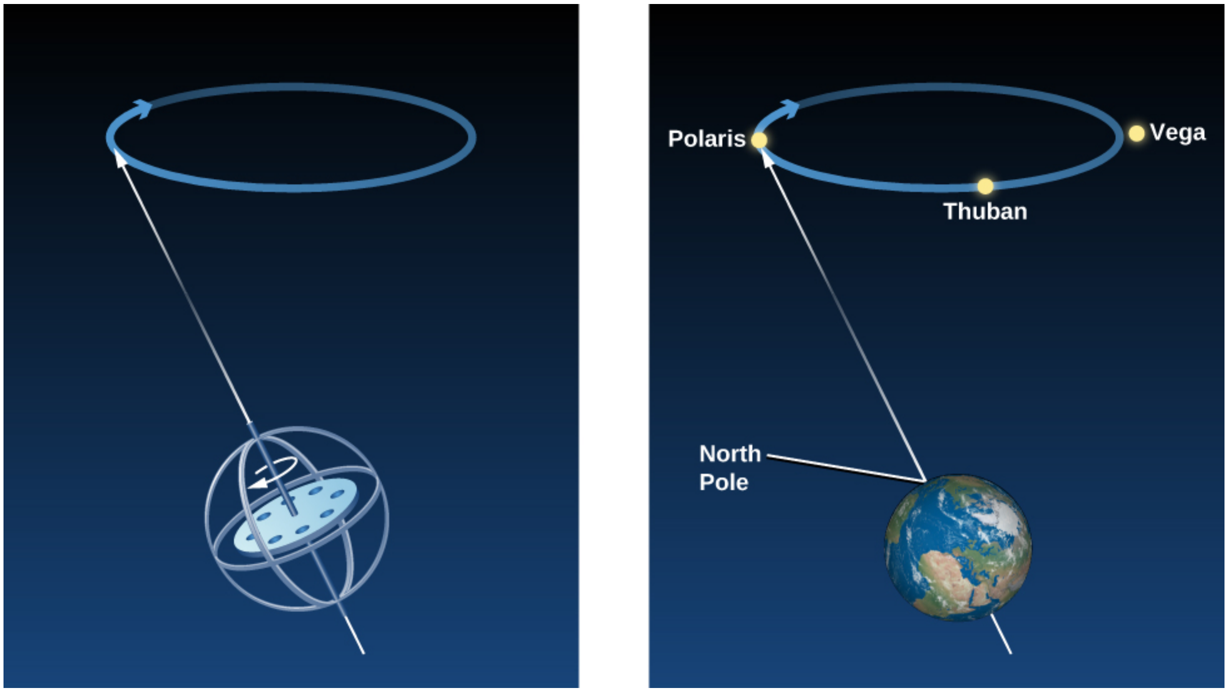
Perhaps the greatest astronomer of antiquity was Hipparchus, born in Nicaea in what is present-day Turkey. He erected an observatory on the island of Rhodes around 150 BCE, when the Roman Republic was expanding its influence throughout the Mediterranean region. There he measured, as accurately as possible, the positions of objects in the sky, compiling a pioneering star catalog with about 850 entries. He designated celestial coordinates for each star, specifying its position in the sky, just as we specify the position of a point on Earth by giving its latitude and longitude.

He also divided the stars into **apparent magnitudes** according to their apparent brightness. He called the brightest ones “stars of the first magnitude”; the next brightest group, “stars of the second magnitude”; and so forth. This rather arbitrary system, in modified form, still remains in use today (although it is less and less useful for professional astronomers).

By observing the stars and comparing his data with older observations, Hipparchus made one of his most remarkable discoveries: the position in the sky of the north celestial pole had altered over the previous century and a half. Hipparchus deduced correctly that this had happened not only during the period covered by his observations, but was in fact happening all the

time: the direction around which the sky appears to rotate changes slowly but continuously. Recall from the section on celestial poles and the celestial equator that the north celestial pole is just the projection of Earth's North Pole into the sky. If the north celestial pole is wobbling around, then Earth itself must be doing the wobbling. Today, we understand that the direction in which Earth's axis points does indeed change slowly but regularly—a motion we call **precession**. If you have ever watched a spinning top wobble, you observed a similar kind of motion. The top's axis describes a path in the shape of a cone, as Earth's gravity tries to topple it ([link]).

Precession.



Just as the axis of a rapidly spinning top wobbles slowly in a circle, so the axis of Earth wobbles in a 26,000-year cycle. Today the north celestial pole is near the star Polaris, but about 5000 years ago it was close to a star called Thuban, and in 14,000 years it will be closest to the star Vega.

Because our planet is not an exact sphere, but bulges a bit at the equator, the pulls of the Sun and Moon cause it to wobble like a top. It takes about 26,000 years for Earth's axis to complete one circle of precession. As a

result of this motion, the point where our axis points in the sky changes as time goes on. While Polaris is the star closest to the north celestial pole today (it will reach its closest point around the year 2100), the star Vega in the constellation of Lyra will be the North Star in 14,000 years.

Ptolemy's Model of the Solar System

The last great astronomer of the Roman era was Claudius Ptolemy (or Ptolemaeus), who flourished in Alexandria in about the year 140. He wrote a mammoth compilation of astronomical knowledge, which today is called by its Arabic name, *Almagest* (meaning "The Greatest"). *Almagest* does not deal exclusively with Ptolemy's own work; it includes a discussion of the astronomical achievements of the past, principally those of Hipparchus. Today, it is our main source of information about the work of Hipparchus and other Greek astronomers.

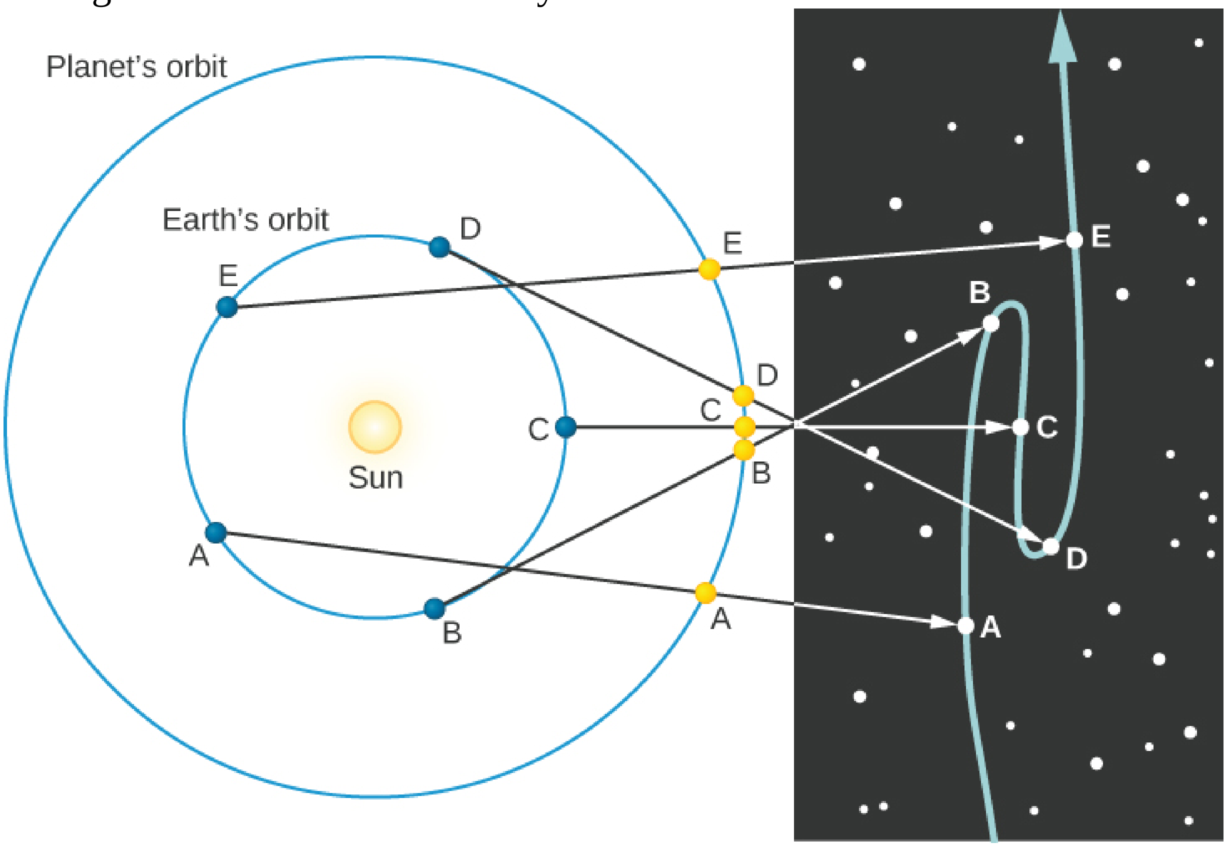
Ptolemy's most important contribution was a geometric representation of the solar system that predicted the positions of the planets for any desired date and time. Hipparchus, not having enough data on hand to solve the problem himself, had instead amassed observational material for posterity to use. Ptolemy supplemented this material with new observations of his own and produced a cosmological model that endured more than a thousand years, until the time of Copernicus.

The complicating factor in explaining the motions of the planets is that their apparent wandering in the sky results from the combination of their own motions with Earth's orbital revolution. As we watch the planets from our vantage point on the moving Earth, it is a little like watching a car race while you are competing in it. Sometimes opponents' cars pass you, but at other times you pass them, making them appear to move backward for a while with respect to you.

[link] shows the motion of Earth and a planet farther from the Sun—in this case, Mars. Earth travels around the Sun in the same direction as the other planet and in nearly the same plane, but its orbital speed is faster. As a result, it overtakes the planet periodically, like a faster race car on the inside track. The figure shows where we see the planet in the sky at different

times. The path of the planet among the stars is illustrated in the star field on the right side of the figure.

Retrograde Motion of a Planet beyond Earth’s Orbit.



The letters on the diagram show where Earth and Mars are at different times. By following the lines from each Earth position through each corresponding Mars position, you can see how the retrograde path of Mars looks against the background stars.

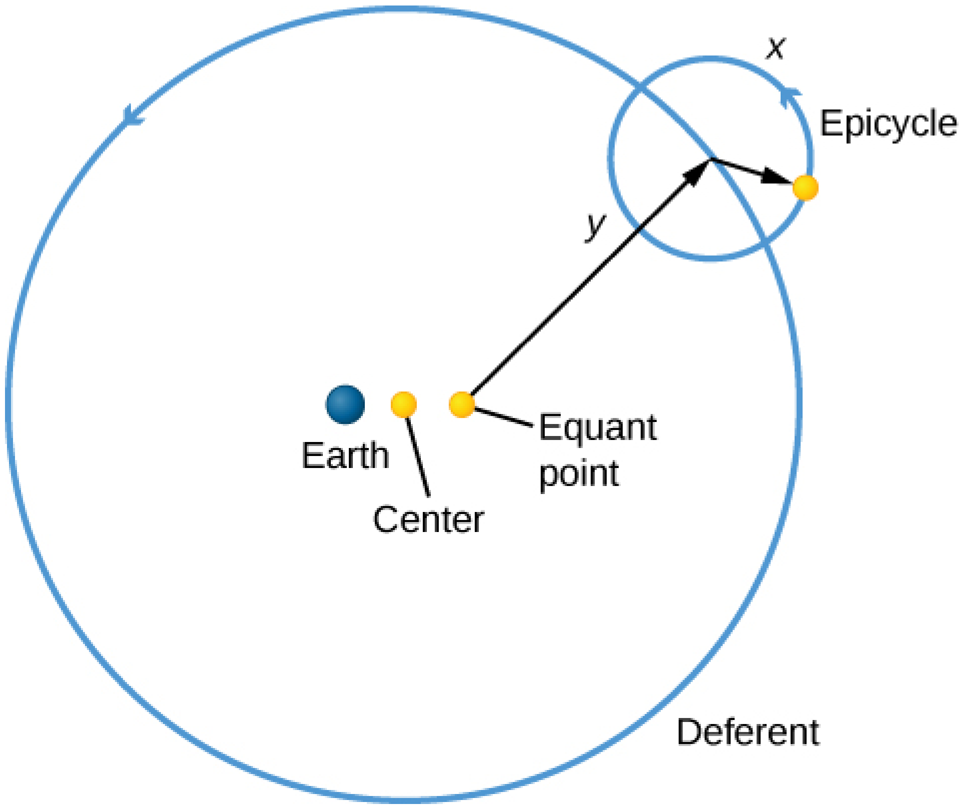
Note:
This retrograde simulation of Mars illustrates the motion of Mars as seen from Earth as well as Earth’s retrograde motion as seen from Mars. There is also an animation of the movement of the two planets relative to each other that creates the appearance of this motion.

Normally, planets move eastward in the sky over the weeks and months as they orbit the Sun, but from positions B to D in [link], as Earth passes the planets in our example, it appears to drift backward, moving west in the sky. Even though it is actually moving to the east, the faster-moving Earth has overtaken it and seems, from our perspective, to be leaving it behind. As Earth rounds its orbit toward position E, the planet again takes up its apparent eastward motion in the sky. The temporary apparent westward motion of a planet as Earth swings between it and the Sun is called **retrograde motion**. Such backward motion is much easier for us to understand today, now that we know Earth is one of the moving planets and not the unmoving center of all creation. But Ptolemy was faced with the far more complex problem of explaining such motion while assuming a stationary Earth.

Furthermore, because the Greeks believed that celestial motions had to be circles, Ptolemy had to construct his model using circles alone. To do it, he needed dozens of circles, some moving around other circles, in a complex structure that makes a modern viewer dizzy. But we must not let our modern judgment cloud our admiration for Ptolemy's achievement. In his day, a complex universe centered on Earth was perfectly reasonable and, in its own way, quite beautiful. However, as Alfonso X, the King of Castile, was reported to have said after having the Ptolemaic system of planet motions explained to him, "If the Lord Almighty had consulted me before embarking upon Creation, I should have recommended something simpler."

Ptolemy solved the problem of explaining the observed motions of planets by having each planet revolve in a small orbit called an **epicycle**. The center of the epicycle then revolved about Earth on a circle called a *deferent* ([link]). When the planet is at position *x* in [link] on the epicycle orbit, it is moving in the same direction as the center of the epicycle; from Earth, the planet appears to be moving eastward. When the planet is at *y*, however, its motion is in the direction opposite to the motion of the epicycle's center around Earth. By choosing the right combination of speeds and distances, Ptolemy succeeded in having the planet moving westward at the correct speed and for the correct interval of time, thus replicating retrograde motion with his model.

Ptolemy's Complicated Cosmological System.



Each planet orbits around a small circle called an *epicycle*. Each epicycle orbits on a larger circle called the *deferent*. This system is not centered exactly on Earth but on an offset point called the *equant*. The Greeks needed all this complexity to explain the actual motions in the sky because they believed that Earth was stationary and that all sky motions had to be circular.

However, we shall see in Orbits and Gravity that the planets, like Earth, travel about the Sun in orbits that are ellipses, not circles. Their actual behavior cannot be represented accurately by a scheme of uniform circular motions. In order to match the observed motions of the planets, Ptolemy had to center the deferent circles, not on Earth, but at points some distance from Earth. In addition, he introduced uniform circular motion around yet another axis, called the *equant point*. All of these considerably complicated his scheme.

It is a tribute to the genius of Ptolemy as a mathematician that he was able to develop such a complex system to account successfully for the observations of planets. It may be that Ptolemy did not intend for his cosmological model to describe reality, but merely to serve as a mathematical representation that allowed him to predict the positions of the planets at any time. Whatever his thinking, his model, with some modifications, was eventually accepted as authoritative in the Muslim world and (later) in Christian Europe.

Ancient Greeks such as Aristotle recognized that Earth and the Moon are spheres, and understood the phases of the Moon, but because of their inability to detect stellar parallax, they rejected the idea that Earth moves. Eratosthenes measured the size of Earth with surprising precision. Hipparchus carried out many astronomical observations, making a star catalog, defining the system of stellar magnitudes, and discovering precession from the apparent shift in the position of the north celestial pole. Ptolemy of Alexandria summarized classic astronomy in his *Almagest*; he explained planetary motions, including retrograde motion, with remarkably good accuracy using a model centered on Earth. This geocentric model, based on combinations of uniform circular motion using epicycles, was accepted as authority for more than a thousand years.

Glossary

apparent magnitude
: a measure of how bright a star looks in the sky; the larger the number, the dimmer the star appears to us

cosmology
: the study of the organization and evolution of the universe

epicycle
: the circular orbit of a body in the Ptolemaic system, the center of which revolves about another circle (the deferent)

parallax

the apparent displacement of a nearby star that results from the motion of Earth around the Sun

precession (of Earth)
: the slow, conical motion of Earth’s axis of rotation caused principally by the gravitational pull of the Moon and Sun on Earth’s equatorial bulge

retrograde motion
: the apparent westward motion of a planet on the celestial sphere or with respect to the stars

The Birth of Modern Astronomy

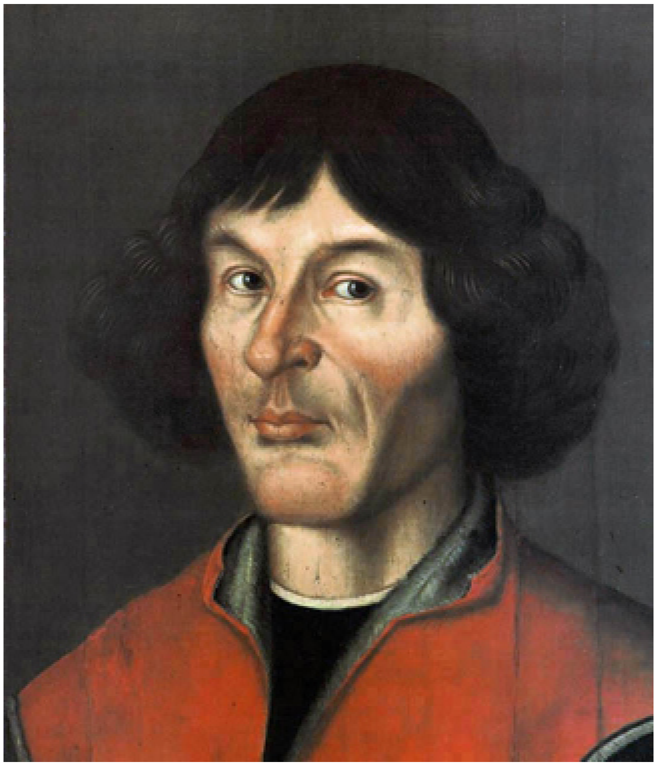
LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how Copernicus developed the heliocentric model of the solar system
- Explain the Copernican model of planetary motion and describe evidence or arguments in favor of it
- Describe Galileo's discoveries concerning the study of motion and forces
- Explain how Galileo's discoveries tilted the balance of evidence in favor of the Copernican model

Astronomy made no major advances in strife-torn medieval Europe. The birth and expansion of Islam after the seventh century led to a flowering of Arabic and Jewish cultures that preserved, translated, and added to many of the astronomical ideas of the Greeks. Many of the names of the brightest stars, for example, are today taken from the Arabic, as are such astronomical terms as "zenith."

As European culture began to emerge from its long, dark age, trading with Arab countries led to a rediscovery of ancient texts such as *Almagest* and to a reawakening of interest in astronomical questions. This time of rebirth (in French, "*renaissance*") in astronomy was embodied in the work of Copernicus ([link]).
Nicolaus Copernicus (1473–1543).



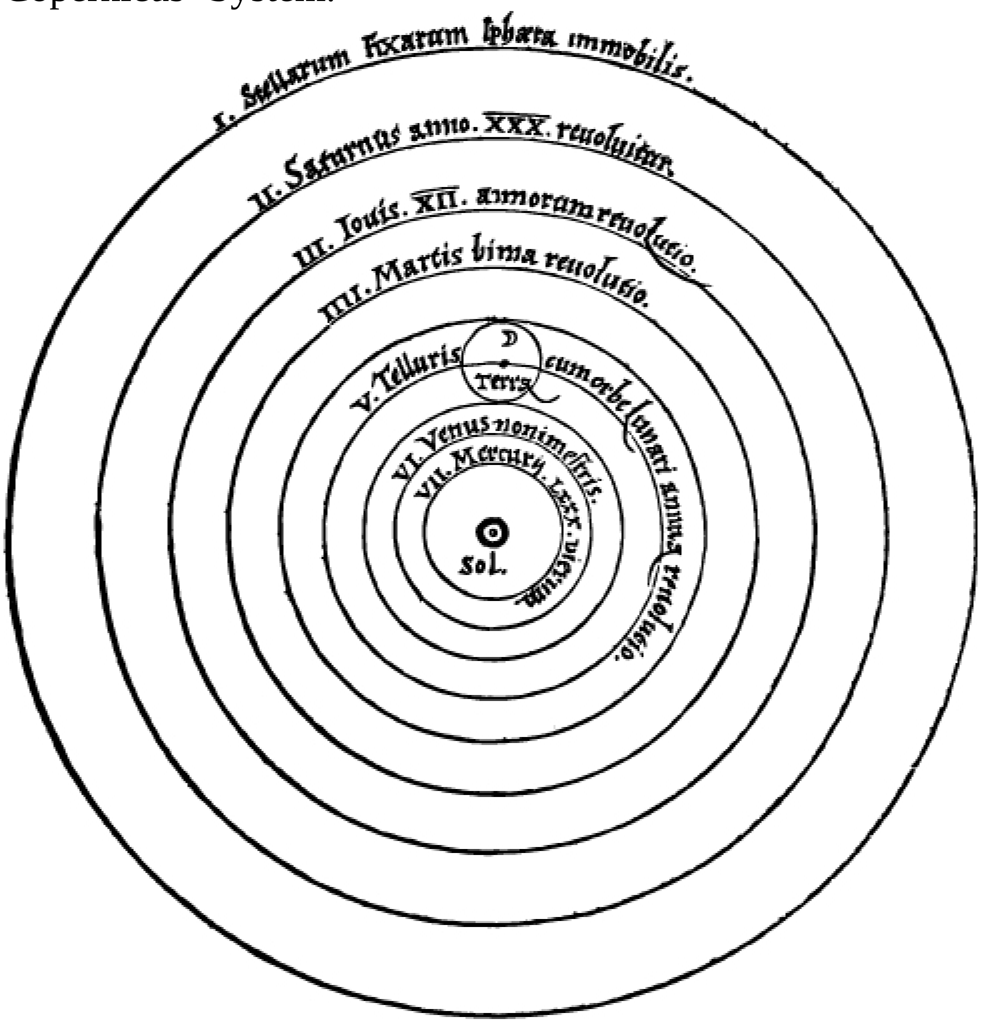
Copernicus was a cleric and scientist who played a leading role in the emergence of modern science. Although he could not prove that Earth revolves about the Sun, he presented such compelling arguments for this idea that he turned the tide of cosmological thought and laid the foundations upon which Galileo and Kepler so effectively built in the following century.

Copernicus

One of the most important events of the Renaissance was the displacement of Earth from the center of the universe, an intellectual revolution initiated

by a Polish cleric in the sixteenth century. Nicolaus Copernicus was born in Torun, a mercantile town along the Vistula River. His training was in law and medicine, but his main interests were astronomy and mathematics. His great contribution to science was a critical reappraisal of the existing theories of planetary motion and the development of a new Sun-centered, or **heliocentric**, model of the solar system. Copernicus concluded that Earth is a planet and that all the planets circle the Sun. Only the Moon orbits Earth ([link]).

Copernicus' System.



Copernicus developed a heliocentric plan of the solar system. This system was published in the first edition of *De Revolutionibus Orbium Coelestium*. Notice the word *Sol* for "Sun" in the middle. (credit: Nicolai Copernici)

Copernicus described his ideas in detail in his book *De Revolutionibus Orbium Coelestium* (*On the Revolution of Celestial Orbs*), published in 1543, the year of his death. By this time, the old Ptolemaic system needed significant adjustments to predict the positions of the planets correctly. Copernicus wanted to develop an improved theory from which to calculate planetary positions, but in doing so, he was himself not free of all traditional prejudices.

He began with several assumptions that were common in his time, such as the idea that the motions of the heavenly bodies must be made up of combinations of uniform circular motions. But he did not assume (as most people did) that Earth had to be in the center of the universe, and he presented a defense of the heliocentric system that was elegant and persuasive. His ideas, although not widely accepted until more than a century after his death, were much discussed among scholars and, ultimately, had a profound influence on the course of world history.

One of the objections raised to the heliocentric theory was that if Earth were moving, we would all sense or feel this motion. Solid objects would be ripped from the surface, a ball dropped from a great height would not strike the ground directly below it, and so forth. But a moving person is not necessarily aware of that motion. We have all experienced seeing an adjacent train, bus, or ship appear to move, only to discover that it is we who are moving.

Copernicus argued that the apparent motion of the Sun about Earth during the course of a year could be represented equally well by a motion of Earth about the Sun. He also reasoned that the apparent rotation of the celestial sphere could be explained by assuming that Earth rotates while the celestial sphere is stationary. To the objection that if Earth rotated about an axis it would fly into pieces, Copernicus answered that if such motion would tear Earth apart, the still faster motion of the much larger celestial sphere required by the geocentric hypothesis would be even more devastating.

The Heliocentric Model

The most important idea in Copernicus' *De Revolutionibus* is that Earth is one of six (then-known) planets that revolve about the Sun. Using this concept, he was able to work out the correct general picture of the solar system. He placed the planets, starting nearest the Sun, in the correct order: Mercury, Venus, Earth, Mars, Jupiter, and Saturn. Further, he deduced that the nearer a planet is to the Sun, the greater its orbital speed. With his theory, he was able to explain the complex retrograde motions of the planets without epicycles and to work out a roughly correct scale for the solar system.

Copernicus could not prove that Earth revolves about the Sun. In fact, with some adjustments, the old Ptolemaic system could have accounted, as well, for the motions of the planets in the sky. But Copernicus pointed out that the Ptolemaic cosmology was clumsy and lacking the beauty and symmetry of its successor.

In Copernicus' time, in fact, few people thought there were ways to prove whether the heliocentric or the older geocentric system was correct. A long philosophical tradition, going back to the Greeks and defended by the Catholic Church, held that pure human thought combined with divine revelation represented the path to truth. Nature, as revealed by our senses, was suspect. For example, Aristotle had reasoned that heavier objects (having more of the quality that made them heavy) must fall to Earth faster than lighter ones. This is absolutely incorrect, as any simple experiment dropping two balls of different weights shows. However, in Copernicus' day, experiments did not carry much weight (if you will pardon the expression); Aristotle's reasoning was more convincing.

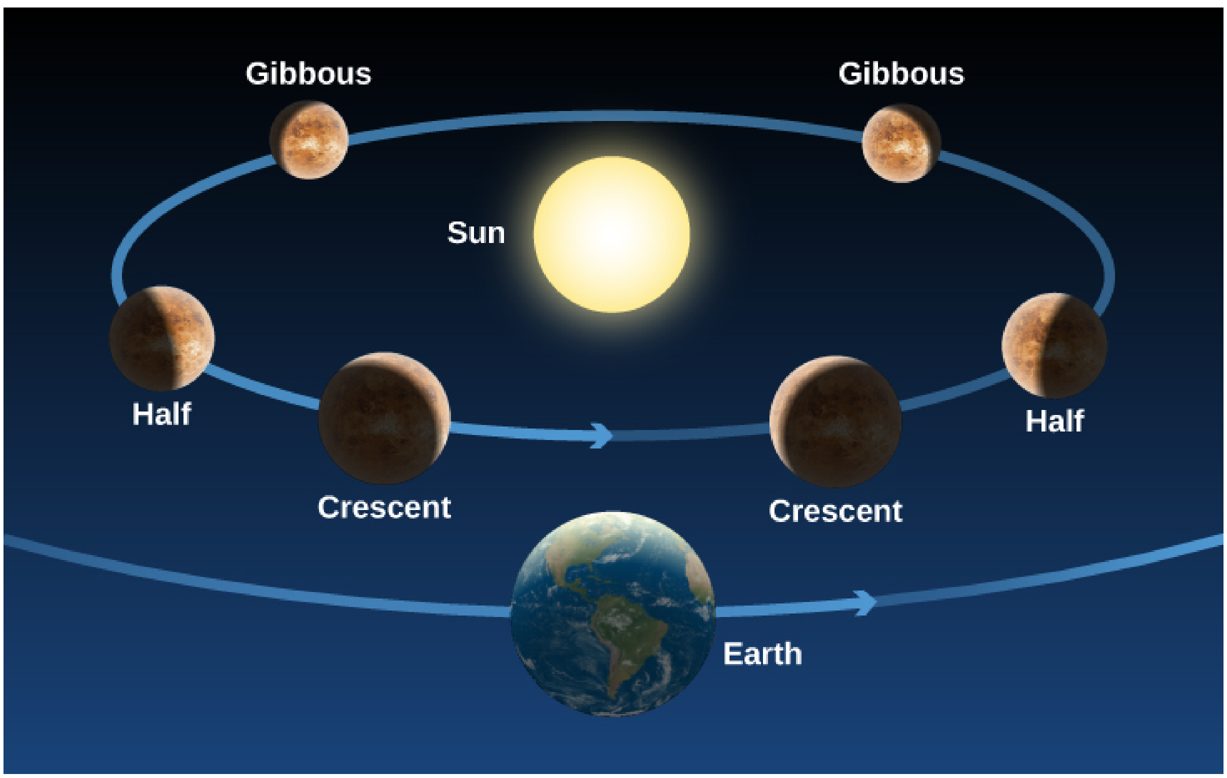
In this environment, there was little motivation to carry out observations or experiments to distinguish between competing cosmological theories (or anything else). It should not surprise us, therefore, that the heliocentric idea was debated for more than half a century without any tests being applied to determine its validity. (In fact, in the North American colonies, the older geocentric system was still taught at Harvard University in the first years after it was founded in 1636.)

Contrast this with the situation today, when scientists rush to test each new hypothesis and do not accept any ideas until the results are in. For example,

when two researchers at the University of Utah announced in 1989 that they had discovered a way to achieve nuclear fusion (the process that powers the stars) at room temperature, other scientists at more than 25 laboratories around the United States attempted to duplicate "cold fusion" within a few weeks—without success, as it turned out. The cold fusion theory soon went down in flames.

How would we look at Copernicus' model today? When a new hypothesis or theory is proposed in science, it must first be checked for consistency with what is already known. Copernicus' heliocentric idea passes this test, for it allows planetary positions to be calculated at least as well as does the geocentric theory. The next step is to determine which predictions the new hypothesis makes that differ from those of competing ideas. In the case of Copernicus, one example is the prediction that, if Venus circles the Sun, the planet should go through the full range of phases just as the Moon does, whereas if it circles Earth, it should not ([link]). Also, we should not be able to see the full phase of Venus from Earth because the Sun would then be between Venus and Earth. But in those days, before the telescope, no one imagined testing these predictions.

Phases of Venus.



As Venus moves around the Sun, we see changing illumination of its surface, just as we see the face of the Moon illuminated differently in the course of a month.

Note:
This animation shows the phases of Venus. You can also see its distance from Earth as it orbits the Sun.

Galileo and the Beginning of Modern Science

Many of the modern scientific concepts of observation, experimentation, and the testing of hypotheses through careful quantitative measurements were pioneered by a man who lived nearly a century after Copernicus. Galileo Galilei ([link]), a contemporary of Shakespeare, was born in Pisa. Like Copernicus, he began training for a medical career, but he had little interest in the subject and later switched to mathematics. He held faculty positions at the University of Pisa and the University of Padua, and eventually became mathematician to the Grand Duke of Tuscany in Florence.

Galileo Galilei (1564–1642).

Galileo advocated that we perform experiments or make observations to ask nature its ways. When Galileo turned the telescope to the sky, he found things were not the way philosophers had supposed.

Galileo’s greatest contributions were in the field of mechanics, the study of motion and the actions of forces on bodies. It was familiar to all persons then, as it is to us now, that if something is at rest, it tends to remain at rest and requires some outside influence to start it in motion. Rest was thus generally regarded as the natural state of matter. Galileo showed, however, that rest is no more natural than motion.

If an object is slid along a rough horizontal floor, it soon comes to rest because friction between it and the floor acts as a retarding force. However,

if the floor and the object are both highly polished, the object, given the same initial speed, will slide farther before stopping. On a smooth layer of ice, it will slide farther still. Galileo reasoned that if all resisting effects could be removed, the object would continue in a steady state of motion indefinitely. He argued that a force is required not only to start an object moving from rest but also to slow down, stop, speed up, or change the direction of a moving object. You will appreciate this if you have ever tried to stop a rolling car by leaning against it, or a moving boat by tugging on a line.

Galileo also studied the way objects **accelerate**—change their speed or direction of motion. Galileo watched objects as they fell freely or rolled down a ramp. He found that such objects accelerate uniformly; that is, in equal intervals of time they gain equal increments in speed. Galileo formulated these newly found laws in precise mathematical terms that enabled future experimenters to predict how far and how fast objects would move in various lengths of time.

Note:
In theory, if Galileo is right, a feather and a hammer, dropped at the same time from a height, should land at the same moment. On Earth, this experiment is not possible because air resistance and air movements make the feather flutter, instead of falling straight down, accelerated only by the force of gravity. For generations, physics teachers had said that the place to try this experiment is somewhere where there is no air, such as the Moon. In 1971, *Apollo 15* astronaut David Scott took a hammer and feather to the Moon and tried it, to the delight of physics nerds everywhere. NASA provides the video of the hammer and feather as well as a brief explanation.

Sometime in the 1590s, Galileo adopted the Copernican hypothesis of a heliocentric solar system. In Roman Catholic Italy, this was not a popular philosophy, for Church authorities still upheld the ideas of Aristotle and Ptolemy, and they had powerful political and economic reasons for insisting

that Earth was the center of creation. Galileo not only challenged this thinking but also had the audacity to write in Italian rather than scholarly Latin, and to lecture publicly on those topics. For him, there was no contradiction between the authority of the Church in matters of religion and morality, and the authority of nature (revealed by experiments) in matters of science. It was primarily because of Galileo and his “dangerous” opinions that, in 1616, the Church issued a prohibition decree stating that the Copernican doctrine was “false and absurd” and not to be held or defended.

Galileo’s Astronomical Observations

It is not certain who first conceived of the idea of combining two or more pieces of glass to produce an instrument that enlarged images of distant objects, making them appear nearer. The first such “spyglasses” (now called *telescopes*) that attracted much notice were made in 1608 by the Dutch spectacle maker Hans Lippershey (1570–1619). Galileo heard of the discovery and, without ever having seen an assembled telescope, constructed one of his own with a three-power magnification (3×), which made distant objects appear three times nearer and larger ([link]).

Telescope Used by Galileo.

The telescope has a wooden tube covered with paper and a lens 26 millimeters across.

On August 25, 1609, Galileo demonstrated a telescope with a magnification of 9× to government officials of the city-state of Venice. By a magnification of 9×, we mean the linear dimensions of the objects being viewed appeared nine times larger or, alternatively, the objects appeared nine times closer than they really were. There were obvious military advantages associated with a device for seeing distant objects. For his invention, Galileo’s salary was nearly doubled, and he was granted lifetime tenure as a professor. (His

university colleagues were outraged, particularly because the invention was not even original.)

Others had used the telescope before Galileo to observe things on Earth. But in a flash of insight that changed the history of astronomy, Galileo realized that he could turn the power of the telescope toward the heavens. Before using his telescope for astronomical observations, Galileo had to devise a stable mount and improve the optics. He increased the magnification to 30×. Galileo also needed to acquire confidence in the telescope.

At that time, human eyes were believed to be the final arbiter of truth about size, shape, and color. Lenses, mirrors, and prisms were known to distort distant images by enlarging, reducing, or inverting them, or spreading the light into a spectrum (rainbow of colors). Galileo undertook repeated experiments to convince himself that what he saw through the telescope was identical to what he saw up close. Only then could he begin to believe that the miraculous phenomena the telescope revealed in the heavens were real.

Beginning his astronomical work late in 1609, Galileo found that many stars too faint to be seen with the unaided eye became visible with his telescope. In particular, he found that some nebulous blurs resolved into many stars, and that the Milky Way—the strip of whiteness across the night sky—was also made up of a multitude of individual stars.

Examining the planets, Galileo found four moons revolving about Jupiter in times ranging from just under 2 days to about 17 days. This discovery was particularly important because it showed that not everything has to revolve around Earth. Furthermore, it demonstrated that there could be centers of motion that are themselves in motion. Defenders of the geocentric view had argued that if Earth was in motion, then the Moon would be left behind because it could hardly keep up with a rapidly moving planet. Yet, here were Jupiter’s moons doing exactly that. (To recognize this discovery and honor his work, NASA named a spacecraft that explored the Jupiter system Galileo.)

With his telescope, Galileo was able to carry out the test of the Copernican theory mentioned earlier, based on the phases of Venus. Within a few months, he had found that Venus goes through phases like the Moon, showing that it must revolve about the Sun, so that we see different parts of its daylight side at different times (see [link].) These observations could not be reconciled with Ptolemy's model, in which Venus circled about Earth. In Ptolemy's model, Venus could also show phases, but they were the wrong phases in the wrong order from what Galileo observed.

Galileo also observed the Moon and saw craters, mountain ranges, valleys, and flat, dark areas that he thought might be water. These discoveries showed that the Moon might be not so dissimilar to Earth—suggesting that Earth, too, could belong to the realm of celestial bodies.

Note:
For more information about the life and work of Galileo, see the Galileo Project at Rice University.

After Galileo's work, it became increasingly difficult to deny the Copernican view, and Earth was slowly dethroned from its central position in the universe and given its rightful place as one of the planets attending the Sun. Initially, however, Galileo met with a great deal of opposition. The Roman Catholic Church, still reeling from the Protestant Reformation, was looking to assert its authority and chose to make an example of Galileo. He had to appear before the Inquisition to answer charges that his work was heretical, and he was ultimately condemned to house arrest. His books were on the Church's forbidden list until 1836, although in countries where the Roman Catholic Church held less sway, they were widely read and discussed. Not until 1992 did the Catholic Church admit publicly that it had erred in the matter of censoring Galileo's ideas.

The new ideas of Copernicus and Galileo began a revolution in our conception of the cosmos. It eventually became evident that the universe is a vast place and that Earth's role in it is relatively unimportant. The idea

that Earth moves around the Sun like the other planets raised the possibility that they might be worlds themselves, perhaps even supporting life. As Earth was demoted from its position at the center of the universe, so, too, was humanity. The universe, despite what we may wish, does not revolve around us.

Most of us take these things for granted today, but four centuries ago such concepts were frightening and heretical for some, immensely stimulating for others. The pioneers of the Renaissance started the European world along the path toward science and technology that we still tread today. For them, nature was rational and ultimately knowable, and experiments and observations provided the means to reveal its secrets.

Note:

Observing the Planets

At most any time of the night, and at any season, you can spot one or more bright planets in the sky. All five of the planets known to the ancients—Mercury, Venus, Mars, Jupiter, and Saturn—are more prominent than any but the brightest stars, and they can be seen even from urban locations if you know where and when to look. One way to tell planets from bright stars is that planets twinkle less.

Venus, which stays close to the Sun from our perspective, appears either as an “evening star” in the west after sunset or as a “morning star” in the east before sunrise. It is the brightest object in the sky after the Sun and Moon. It far outshines any real star, and under the most favorable circumstances, it can even cast a visible shadow. Some young military recruits have tried to shoot Venus down as an approaching enemy craft or UFO.

Mars, with its distinctive red color, can be nearly as bright as Venus is when close to Earth, but normally it remains much less conspicuous. Jupiter is most often the second-brightest planet, approximately equaling in brilliance the brightest stars. Saturn is dimmer, and it varies considerably in brightness, depending on whether its large rings are seen nearly edge-on (faint) or more widely opened (bright).

Mercury is quite bright, but few people ever notice it because it never moves very far from the Sun (it’s never more than 28° away in the sky) and

is always seen against bright twilight skies.
True to their name, the planets “wander” against the background of the “fixed” stars. Although their apparent motions are complex, they reflect an underlying order upon which the heliocentric model of the solar system, as described in this chapter, was based. The positions of the planets are often listed in newspapers (sometimes on the weather page), and clear maps and guides to their locations can be found each month in such magazines as *Sky & Telescope* and *Astronomy* (available at most libraries and online). There are also a number of computer programs and phone and tablet apps that allow you to display where the planets are on any night.

Nicolaus Copernicus introduced the heliocentric cosmology to Renaissance Europe in his book *De Revolutionibus*. Although he retained the Aristotelian idea of uniform circular motion, Copernicus suggested that Earth is a planet and that the planets all circle about the Sun, dethroning Earth from its position at the center of the universe. Galileo was the father of both modern experimental physics and telescopic astronomy. He studied the acceleration of moving objects and, in 1610, began telescopic observations, discovering the nature of the Milky Way, the large-scale features of the Moon, the phases of Venus, and four moons of Jupiter. Although he was accused of heresy for his support of heliocentric cosmology, Galileo is credited with observations and brilliant writings that convinced most of his scientific contemporaries of the reality of the Copernican theory.

For Further Exploration

Articles

Ancient Astronomy

Gingerich, O. “From Aristarchus to Copernicus.” *Sky & Telescope* (November 1983): 410.

Gingerich, O. “Islamic Astronomy.” *Scientific American* (April 1986): 74.

Astronomy and Astrology

Fraknoi, A. “Your Astrology Defense Kit.” *Sky & Telescope* (August 1989): 146.

Copernicus and Galileo

Gingerich, O. “Galileo and the Phases of Venus.” *Sky & Telescope* (December 1984): 520.

Gingerich, O. “How Galileo Changed the Rules of Science.” *Sky & Telescope* (March 1993): 32.

Maran, S., and Marschall, L. “The Moon, the Telescope, and the Birth of the Modern World.” *Sky & Telescope* (February 2009): 28.

Sobel, D. “The Heretic’s Daughter: A Startling Correspondence Reveals a New Portrait of Galileo.” *The New Yorker* (September 13, 1999): 52.

Websites

Ancient Astronomy

Aristarchos of Samos: http://adsabs.harvard.edu//full/seri/JRASC/0075//0000029.000.html. By Dr. Alan Batten.

Claudius Ptolemy: http://www-history.mcs.st-and.ac.uk/Biographies/Ptolemy.html. An interesting biography.

Hipparchus of Rhodes: http://www-history.mcs.st-andrews.ac.uk/Biographies/Hipparchus.html. An interesting biography.

Astronomy and Astrology

Astrology and Science: http://www.astrology-and-science.com/hpage.htm. The best site for a serious examination of the issues with astrology and the research on whether it works.

Real Romance in the Stars: http://www.independent.co.uk/voices/the-real-romance-in-the-stars-1527970.html. 1995 newspaper commentary attacking astrology.

Copernicus and Galileo

Galileo Galilei: http://www-history.mcs.st-andrews.ac.uk/Biographies/Galileo.html. A good biography with additional links.

Galileo Project: http://galileo.rice.edu/. Rice University's repository of information on Galileo.

Nicolaus Copernicus: http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Copernicus.html. A biography including links to photos about his life.

Videos

Astronomy and Astrology

Astrology Debunked: https://www.youtube.com/watch?v=y84HX2pMo5U. A compilation of scientists and magicians commenting skeptically on

astrology (9:09).

Copernicus and Galileo

Galileo: http://www.biography.com/people/galileo-9305220. A brief biography (2:51).

Galileo’s Battle for the Heavens: https://www.youtube.com/watch?v=jvlr2iMWQyc. A NOVA episode on PBS (1:48:55)

Nicolaus Copernicus: http://www.biography.com/people/nicolaus-copernicus-9256984. An overview of his life and work (2:41).

Collaborative Group Activities

A. With your group, consider the question with which we began this chapter. How many ways can you think of to prove to a member of the “Flat Earth Society” that our planet is, indeed, round?
B. Make a list of ways in which a belief in astrology (the notion that your life path or personality is controlled by the position of the Sun, Moon, and planets at the time of your birth) might be harmful to an individual or to society at large.
C. Have members of the group compare their experiences with the night sky. Did you see the Milky Way? Can you identify any constellations? Make a list of reasons why you think so many fewer people know the night sky today than at the time of the ancient Greeks. Discuss reasons for why a person, today, may want to be acquainted with the night sky.
D. Constellations commemorate great heroes, dangers, or events in the legends of the people who name them. Suppose we had to start from scratch today, naming the patterns of stars in the sky. Whom or what would you choose to commemorate by naming a constellation after it, him, or her and why (begin with people from history; then if you have time, include living people as well)? Can the members of your group agree on any choices?
E. Although astronomical mythology no longer holds a powerful sway over the modern imagination, we still find proof of the power of

astronomical images in the number of products in the marketplace that have astronomical names. How many can your group come up with? (Think of things like Milky Way candy bars, Eclipse and Orbit gum, or Comet cleanser.)

Review Questions

Exercise:

Problem:

From where on Earth could you observe all of the stars during the course of a year? What fraction of the sky can be seen from the North Pole?

Exercise:

Problem: Give four ways to demonstrate that Earth is spherical.

Exercise:

Problem:

Explain, according to both geocentric and heliocentric cosmologies, why we see retrograde motion of the planets.

Exercise:

Problem:

In what ways did the work of Copernicus and Galileo differ from the views of the ancient Greeks and of their contemporaries?

Exercise:

Problem:

What were four of Galileo’s discoveries that were important to astronomy?

Exercise:

Problem:

Explain the origin of the magnitude designation for determining the brightness of stars. Why does it seem to go backward, with smaller numbers indicating brighter stars?

Exercise:

Problem:

Ursa Minor contains the pole star, Polaris, and the asterism known as the Little Dipper. From most locations in the Northern Hemisphere, all of the stars in Ursa Minor are circumpolar. Does that mean these stars are also above the horizon during the day? Explain.

Exercise:

Problem:

How many degrees does the Sun move per day relative to the fixed stars? How many days does it take for the Sun to return to its original location relative to the fixed stars?

Exercise:

Problem:

How many degrees does the Moon move per day relative to the fixed stars? How many days does it take for the Moon to return to its original location relative to the fixed stars?

Exercise:

Problem:

Explain how the zodiacal constellations are different from the other constellations.

Exercise:

Problem: The Sun was once thought to be a planet. Explain why.

Exercise:

Problem:

Is the ecliptic the same thing as the celestial equator? Explain.

Exercise:

Problem: What is an asterism? Can you name an example?

Exercise:

Problem: Why did Pythagoras believe that Earth should be spherical?

Exercise:

Problem:

How did Aristotle deduce that the Sun is farther away from Earth than the Moon?

Exercise:

Problem:

What are two ways in which Aristotle deduced that Earth is spherical?

Exercise:

Problem:

How did Hipparchus discover the wobble of Earth’s axis, known as *precession*?

Exercise:

Problem:

Why did Ptolemy have to introduce multiple circles of motion for the planets instead of a single, simple circle to represent the planet’s motion around the Earth?

Exercise:

Problem:

Why did Copernicus want to develop a completely new system for predicting planetary positions? Provide two reasons.

Exercise:

Problem:

What two factors made it difficult, at first, for astronomers to choose between the Copernican heliocentric model and the Ptolemaic geocentric model?

Exercise:

Problem:

What phases would Venus show if the geocentric model were correct?

Thought Questions

Exercise:

Problem:

Describe a practical way to determine in which constellation the Sun is found at any time of the year.

Exercise:

Problem:

What is a constellation as astronomers define it today? What does it mean when an astronomer says, “I saw a comet in Orion last night”?

Exercise:

Problem:

Draw a picture that explains why Venus goes through phases the way the Moon does, according to the heliocentric cosmology. Does Jupiter also go through phases as seen from Earth? Why?

Exercise:

Problem:

Show with a simple diagram how the lower parts of a ship disappear first as it sails away from you on a spherical Earth. Use the same diagram to show why lookouts on old sailing ships could see farther from the masthead than from the deck. Would there be any advantage to posting lookouts on the mast if Earth were flat? (Note that these nautical arguments for a spherical Earth were quite familiar to Columbus and other mariners of his time.)

Exercise:

Problem:

Parallaxes of stars were not observed by ancient astronomers. How can this fact be reconciled with the heliocentric hypothesis?

Exercise:

Problem:

Why do you think so many people still believe in astrology and spend money on it? What psychological needs does such a belief system satisfy?

Exercise:

Problem:

Consider three cosmological perspectives—the geocentric perspective, the heliocentric perspective, and the modern perspective—in which the Sun is a minor star on the outskirts of one galaxy among billions. Discuss some of the cultural and philosophical implications of each point of view.

Exercise:

Problem:

The north celestial pole appears at an altitude above the horizon that is equal to the observer’s latitude. Identify Polaris, the North Star, which lies very close to the north celestial pole. Measure its altitude. (This can be done with a protractor. Alternatively, your fist, extended at arm’s length, spans a distance approximately equal to 10°.) Compare this estimate with your latitude. (Note that this experiment cannot be performed easily in the Southern Hemisphere because Polaris itself is not visible in the south and no bright star is located near the south celestial pole.)

Exercise:

Problem:

What were two arguments or lines of evidence in support of the geocentric model?

Exercise:

Problem:

Although the Copernican system was largely correct to place the Sun at the center of all planetary motion, the model still gave inaccurate predictions for planetary positions. Explain the flaw in the Copernican model that hindered its accuracy.

Exercise:

Problem:

During a retrograde loop of Mars, would you expect Mars to be brighter than usual in the sky, about average in brightness, or fainter than usual in the sky? Explain.

Exercise:

Problem:

The Great Pyramid of Giza was constructed nearly 5000 years ago. Within the pyramid, archaeologists discovered a shaft leading from the central chamber out of the pyramid, oriented for favorable viewing of the bright star Thuban at that time. Thinking about Earth's precession, explain why Thuban might have been an important star to the ancient Egyptians.

Exercise:

Problem:

Explain why more stars are circumpolar for observers at higher latitudes.

Exercise:

Problem:

What is the altitude of the north celestial pole in the sky from your latitude? If you do not know your latitude, look it up. If you are in the Southern Hemisphere, answer this question for the south celestial pole, since the north celestial pole is not visible from your location.

Exercise:

Problem:

If you were to drive to some city south of your current location, how would the altitude of the celestial pole in the sky change?

Exercise:

Problem:

Hipparchus could have warned us that the dates associated with each of the natal astrology sun signs would eventually be wrong. Explain why.

Exercise:

Problem:

Explain three lines of evidence that argue against the validity of astrology.

Exercise:

Problem:

What did Galileo discover about the planet Jupiter that cast doubt on exclusive geocentrism?

Exercise:

Problem:

What did Galileo discover about Venus that cast doubt on geocentrism?

Figuring for Yourself

Exercise:

Problem:

Suppose Eratosthenes had found that, in Alexandria, at noon on the first day of summer, the line to the Sun makes an angle 30° with the vertical. What, then, would he have found for Earth's circumference?

Exercise:

Problem:

Suppose Eratosthenes' results for Earth's circumference were quite accurate. If the diameter of Earth is 12,740 km, what is the length of his stadium in kilometers?

Exercise:

Problem:

Suppose you are on a strange planet and observe, at night, that the stars do not rise and set, but circle parallel to the horizon. Next, you walk in a constant direction for 8000 miles, and at your new location on the planet, you find that all stars rise straight up in the east and set straight down in the west, perpendicular to the horizon. How could you determine the circumference of the planet without any further observations? What is the circumference, in miles, of the planet?

Glossary

accelerate
: to change velocity; to speed up, slow down, or change direction.

heliocentric
: centered on the Sun

Thinking Ahead
class="introduction"
International Space Station.

This space habitat and laboratory orbits Earth once every 90 minutes. (credit: modification of work by NASA)

How would you find a new planet at the outskirts of our solar system that is too dim to be seen with the unaided eye and is so far away that it moves very slowly among the stars? This was the problem confronting astronomers during the nineteenth century as they tried to pin down a full inventory of our solar system.

If we could look down on the solar system from somewhere out in space, interpreting planetary motions would be much simpler. But the fact is, we must observe the positions of all the other planets from our own moving

planet. Scientists of the Renaissance did not know the details of Earth's motions any better than the motions of the other planets. Their problem, as we saw in Observing the Sky: The Birth of Astronomy, was that they had to deduce the nature of all planetary motion using only their earthbound observations of the other planets' positions in the sky. To solve this complex problem more fully, better observations and better models of the planetary system were needed.

The Laws of Planetary Motion

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe how Tycho Brahe and Johannes Kepler contributed to our understanding of how planets move around the Sun
- Explain Kepler's three laws of planetary motion

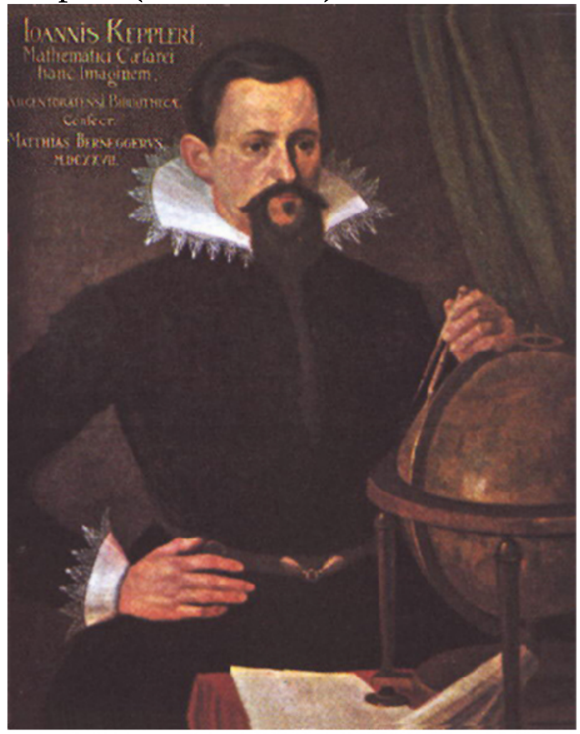
At about the time that Galileo was beginning his experiments with falling bodies, the efforts of two other scientists dramatically advanced our understanding of the motions of the planets. These two astronomers were the observer Tycho Brahe and the mathematician Johannes Kepler. Together, they placed the speculations of Copernicus on a sound mathematical basis and paved the way for the work of Isaac Newton in the next century.

Tycho Brahe's Observatory

Three years after the publication of Copernicus' *De Revolutionibus*, Tycho Brahe was born to a family of Danish nobility. He developed an early interest in astronomy and, as a young man, made significant astronomical observations. Among these was a careful study of what we now know was an exploding star that flared up to great brilliance in the night sky. His growing reputation gained him the patronage of the Danish King Frederick II, and at the age of 30, Brahe was able to establish a fine astronomical observatory on the North Sea island of Hven ([link]). Brahe was the last and greatest of the pre-telescopic observers in Europe.

Tycho Brahe (1546–1601) and Johannes Kepler (1571–1630).

(a)



(b)

(a) A stylized engraving shows Tycho Brahe using his instruments to measure the altitude of celestial objects above the horizon. The large curved instrument in the foreground allowed him to measure precise angles in the sky. Note that the scene includes hints of the grandeur of Brahe's observatory at Hven. (b) Kepler was a German mathematician and astronomer. His discovery of the basic laws that describe planetary motion placed the heliocentric cosmology of Copernicus on a firm mathematical basis.

At Hven, Brahe made a continuous record of the positions of the Sun, Moon, and planets for almost 20 years. His extensive and precise observations enabled him to note that the positions of the planets varied from those given in published tables, which were based on the work of Ptolemy. These data were extremely valuable, but Brahe didn't have the ability to analyze them and develop a better model than what Ptolemy had published. He was further inhibited because he was an extravagant and

cantankerous fellow, and he accumulated enemies among government officials. When his patron, Frederick II, died in 1597, Brahe lost his political base and decided to leave Denmark. He took up residence in Prague, where he became court astronomer to Emperor Rudolf of Bohemia. There, in the year before his death, Brahe found a most able young mathematician, Johannes Kepler, to assist him in analyzing his extensive planetary data.

Johannes Kepler

Johannes Kepler was born into a poor family in the German province of Württemberg and lived much of his life amid the turmoil of the Thirty Years' War (see [link]). He attended university at Tubingen and studied for a theological career. There, he learned the principles of the Copernican system and became converted to the heliocentric hypothesis. Eventually, Kepler went to Prague to serve as an assistant to Brahe, who set him to work trying to find a satisfactory theory of planetary motion—one that was compatible with the long series of observations made at Hven. Brahe was reluctant to provide Kepler with much material at any one time for fear that Kepler would discover the secrets of the universal motion by himself, thereby robbing Brahe of some of the glory. Only after Brahe's death in 1601 did Kepler get full possession of the priceless records. Their study occupied most of Kepler's time for more than 20 years.

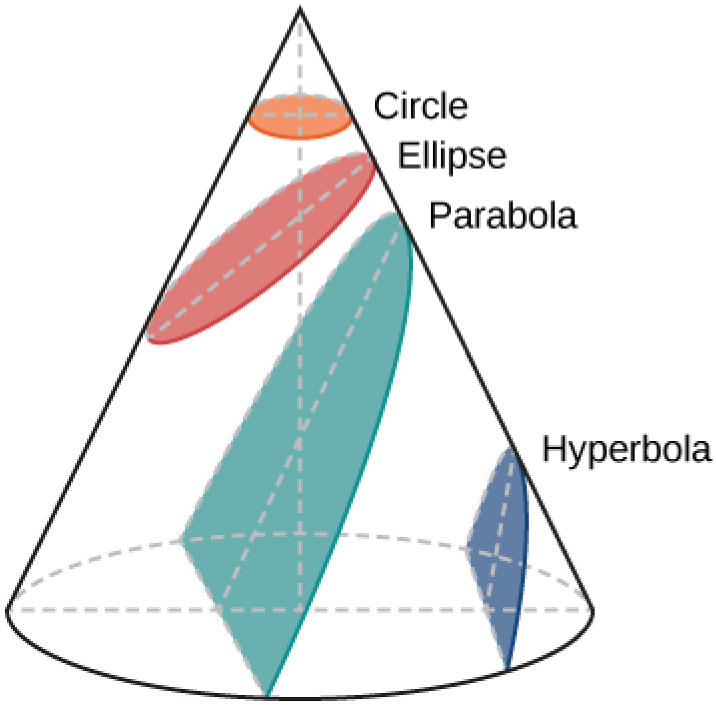
Through his analysis of the motions of the planets, Kepler developed a series of principles, now known as *Kepler's three laws*, which described the behavior of planets based on their paths through space. The first two laws of planetary motion were published in 1609 in *The New Astronomy*. Their discovery was a profound step in the development of modern science.

The First Two Laws of Planetary Motion

The path of an object through space is called its **orbit**. Kepler initially assumed that the orbits of planets were circles, but doing so did not allow him to find orbits that were consistent with Brahe's observations. Working with the data for Mars, he eventually discovered that the orbit of that planet had the shape of a somewhat flattened circle, or **ellipse**. Next to the circle,

the ellipse is the simplest kind of closed curve, belonging to a family of curves known as *conic sections* ([link]).

Conic Sections.



The circle, ellipse, parabola, and hyperbola are all formed by the intersection of a plane with a cone. This is why such curves are called conic sections.

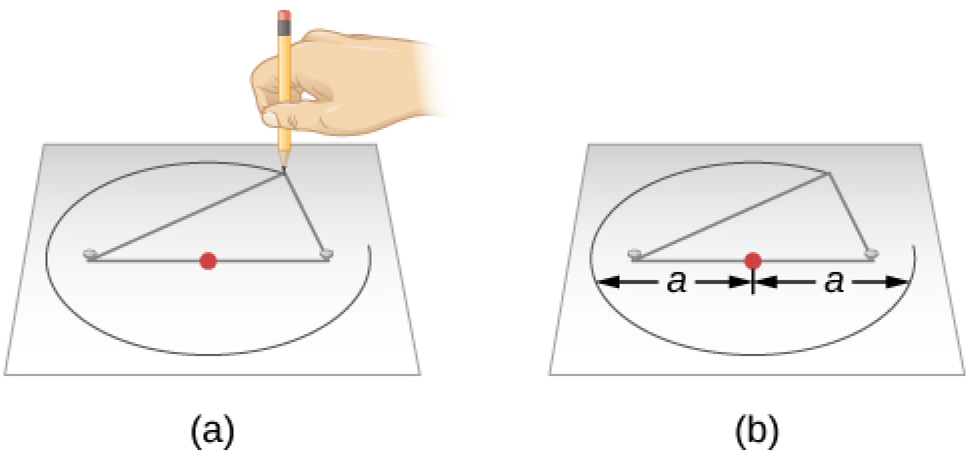
You might recall from math classes that in a circle, the center is a special point. The distance from the center to anywhere on the circle is exactly the same. In an ellipse, the sum of the distance from two special points inside the ellipse to any point on the ellipse is always the same. These two points inside the ellipse are called its foci (singular: **focus**), a word invented for this purpose by Kepler.

This property suggests a simple way to draw an ellipse ([link]). We wrap the ends of a loop of string around two tacks pushed through a sheet of paper into a drawing board, so that the string is slack. If we push a pencil

against the string, making the string taut, and then slide the pencil against the string all around the tacks, the curve that results is an ellipse. At any point where the pencil may be, the sum of the distances from the pencil to the two tacks is a constant length—the length of the string. The tacks are at the two foci of the ellipse.

The widest diameter of the ellipse is called its **major axis**. Half this distance—that is, the distance from the center of the ellipse to one end—is the **semimajor axis**, which is usually used to specify the size of the ellipse. For example, the semimajor axis of the orbit of Mars, which is also the planet's average distance from the Sun, is 228 million kilometers.

Drawing an Ellipse.



(a) We can construct an ellipse by pushing two tacks (the white objects) into a piece of paper on a drawing board, and then looping a string around the tacks. Each tack represents a focus of the ellipse, with one of the tacks being the Sun. Stretch the string tight using a pencil, and then move the pencil around the tacks. The length of the string remains the same, so that the sum of the distances from any point on the ellipse to the foci is always constant. (b) In this illustration, each semimajor axis is denoted by *a*. The distance 2*a* is called the major axis of the ellipse.

The shape (roundness) of an ellipse depends on how close together the two foci are, compared with the major axis. The ratio of the distance between the foci to the length of the semimajor axis is called the **eccentricity** of the ellipse.

If the foci (or tacks) are moved to the same location, then the distance between the foci would be zero. This means that the eccentricity is zero and the ellipse is just a circle; thus, a circle can be called an ellipse of zero eccentricity. In a circle, the semimajor axis would be the radius.

Next, we can make ellipses of various elongations (or extended lengths) by varying the spacing of the tacks (as long as they are not farther apart than the length of the string). The greater the eccentricity, the more elongated is the ellipse, up to a maximum eccentricity of 1.0, when the ellipse becomes "flat," the other extreme from a circle.

The size and shape of an ellipse are completely specified by its semimajor axis and its eccentricity. Using Brahe's data, Kepler found that Mars has an elliptical orbit, with the Sun at one focus (the other focus is empty). The eccentricity of the orbit of Mars is only about 0.1; its orbit, drawn to scale, would be practically indistinguishable from a circle, but the difference turned out to be critical for understanding planetary motions.

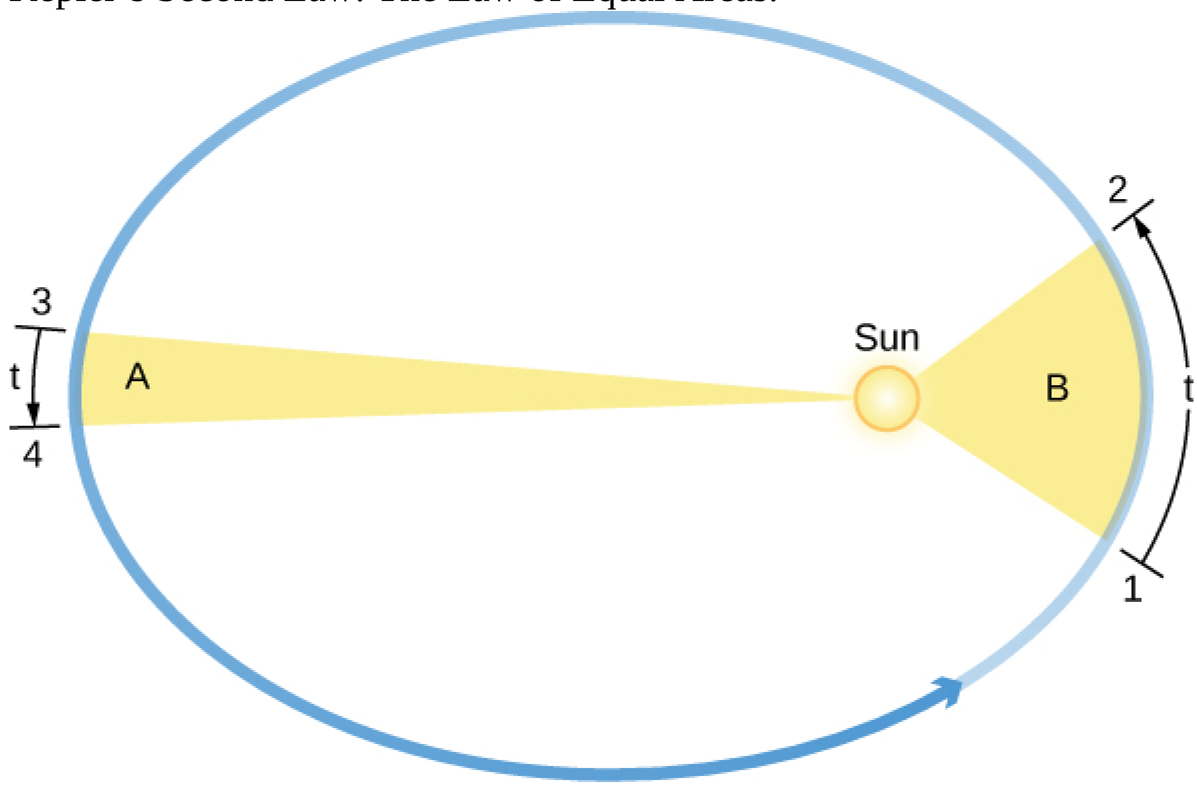
Kepler generalized this result in his first law and said that *the orbits of all the planets are ellipses*. Here was a decisive moment in the history of human thought: it was not necessary to have only circles in order to have an acceptable cosmos. The universe could be a bit more complex than the Greek philosophers had wanted it to be.

Kepler's second law deals with the speed with which each planet moves along its ellipse, also known as its **orbital speed**. Working with Brahe's observations of Mars, Kepler discovered that the planet speeds up as it comes closer to the Sun and slows down as it pulls away from the Sun. He expressed the precise form of this relationship by imagining that the Sun and Mars are connected by a straight, elastic line. When Mars is closer to the Sun (positions 1 and 2 in [link]), the elastic line is not stretched as much, and the planet moves rapidly. Farther from the Sun, as in positions 3 and 4, the line is stretched a lot, and the planet does not move so fast. As

Mars travels in its elliptical orbit around the Sun, the elastic line sweeps out areas of the ellipse as it moves (the colored regions in our figure). Kepler found that in equal intervals of time (t), the areas swept out in space by this imaginary line are always equal; that is, the area of the region B from 1 to 2 is the same as that of region A from 3 to 4.

If a planet moves in a circular orbit, the elastic line is always stretched the same amount and the planet moves at a constant speed around its orbit. But, as Kepler discovered, in most orbits that speed of a planet orbiting its star (or moon orbiting its planet) tends to vary because the orbit is elliptical.

Kepler’s Second Law: The Law of Equal Areas.



The orbital speed of a planet traveling around the Sun (the circular object inside the ellipse) varies in such a way that in equal intervals of time (t), a line between the Sun and a planet sweeps out equal areas (A and B). Note that the eccentricities of the planets’ orbits in our solar system are substantially less than shown here.

Kepler’s Third Law

Kepler’s first two laws of planetary motion describe the shape of a planet’s orbit and allow us to calculate the speed of its motion at any point in the orbit. Kepler was pleased to have discovered such fundamental rules, but they did not satisfy his quest to fully understand planetary motions. He wanted to know why the orbits of the planets were spaced as they are and to find a mathematical pattern in their movements—a “harmony of the spheres” as he called it. For many years he worked to discover mathematical relationships governing planetary spacing and the time each planet took to go around the Sun.

In 1619, Kepler discovered a basic relationship to relate the planets’ orbits to their relative distances from the Sun. We define a planet’s **orbital period**, (***P***), as the time it takes a planet to travel once around the Sun. Also, recall that a planet’s semimajor axis, *a*, is equal to its average distance from the Sun. The relationship, now known as *Kepler’s third law*, says that a planet’s orbital period squared is proportional to the semimajor axis of its orbit cubed, or

Equation:

$$P^2 \propto a^3$$

When *P* (the orbital period) is measured in years, and *a* is expressed in a quantity known as an **astronomical unit (AU)**, the two sides of the formula are not only proportional but equal. One AU is the average distance between Earth and the Sun and is approximately equal to 1.5×10^8 kilometers. In these units,

Equation:

$$P^2 = a^3$$

Kepler’s third law applies to all objects orbiting the Sun, including Earth, and provides a means for calculating their relative distances from the Sun from the time they take to orbit. Let’s look at a specific example to illustrate how useful Kepler’s third law is.

For instance, suppose you time how long Mars takes to go around the Sun (in Earth years). Kepler's third law can then be used to calculate Mars' average distance from the Sun. Mars' orbital period (1.88 Earth years) squared, or P^2, is $1.88^2 = 3.53$, and according to the equation for Kepler's third law, this equals the cube of its semimajor axis, or a^3. So what number must be cubed to give 3.53? The answer is 1.52 (since $1.52 \times 1.52 \times 1.52 = 3.53$). Thus, Mars' semimajor axis in astronomical units must be 1.52 AU. In other words, to go around the Sun in a little less than two years, Mars must be about 50% (half again) as far from the Sun as Earth is.

Example:

Calculating Periods

Imagine an object is traveling around the Sun. What would be the orbital period of the object if its orbit has a semimajor axis of 50 AU?

Solution

From Kepler's third law, we know that (when we use units of years and AU)

Equation:

$$P^2 = a^3$$

If the object's orbit has a semimajor axis of 50 AU (a = 50), we can cube 50 and then take the square root of the result to get P:

Equation:

$$\begin{aligned} P &= \sqrt{a^3} \\ P &= \sqrt{50 \times 50 \times 50} = \sqrt{125{,}000} = 353.6 \text{ years} \end{aligned}$$

Therefore, the orbital period of the object is about 350 years. This would place our hypothetical object beyond the orbit of Pluto.

Check Your Learning

What would be the orbital period of an asteroid (a rocky chunk between Mars and Jupiter) with a semimajor axis of 3 AU?

Note:
Answer:
$P = \sqrt{3\ \times\ 3\ \times\ 3} = \sqrt{27} = 5.2 \text{ years}$

Kepler’s three laws of planetary motion can be summarized as follows:

- **Kepler’s first law**: Each planet moves around the Sun in an orbit that is an ellipse, with the Sun at one focus of the ellipse.
- **Kepler’s second law**: The straight line joining a planet and the Sun sweeps out equal areas in space in equal intervals of time.
- **Kepler’s third law**: The square of a planet’s orbital period is directly proportional to the cube of the semimajor axis of its orbit.

Kepler’s three laws provide a precise geometric description of planetary motion within the framework of the Copernican system. With these tools, it was possible to calculate planetary positions with greatly improved precision. Still, Kepler’s laws are purely descriptive: they do not help us understand what forces of nature constrain the planets to follow this particular set of rules. That step was left to Isaac Newton.

Example:
Applying Kepler’s Third Law
Using the orbital periods and semimajor axes for Venus and Earth that are provided here, calculate P^2 and a^3, and verify that they obey Kepler’s third law. Venus’ orbital period is 0.62 year, and its semimajor axis is 0.72 AU. Earth’s orbital period is 1.00 year, and its semimajor axis is 1.00 AU.
Solution
We can use the equation for Kepler’s third law, $P^2 \propto a^3$. For Venus, $P^2 = 0.62 \times 0.62 = 0.38$ and $a^3 = 0.72 \times 0.72 \times 0.72 = 0.37$ (rounding numbers sometimes causes minor discrepancies like this). The square of the orbital period (0.38) approximates the cube of the semimajor axis (0.37).

Therefore, Venus obeys Kepler's third law. For Earth, $P^2 = 1.00 \times 1.00 = 1.00$ and $a^3 = 1.00 \times 1.00 \times 1.00 = 1.00$. The square of the orbital period (1.00) approximates (in this case, equals) the cube of the semimajor axis (1.00). Therefore, Earth obeys Kepler's third law.

Check Your Learning

Using the orbital periods and semimajor axes for Saturn and Jupiter that are provided here, calculate P^2 and a^3, and verify that they obey Kepler's third law. Saturn's orbital period is 29.46 years, and its semimajor axis is 9.54 AU. Jupiter's orbital period is 11.86 years, and its semimajor axis is 5.20 AU.

Note:

Answer:

For Saturn, $P^2 = 29.46 \times 29.46 = 867.9$ and $a^3 = 9.54 \times 9.54 \times 9.54 = 868.3$. The square of the orbital period (867.9) approximates the cube of the semimajor axis (868.3). Therefore, Saturn obeys Kepler's third law.

Note:

In honor of the scientist who first devised the laws that govern the motions of planets, the team that built the first spacecraft to search for planets orbiting other stars decided to name the probe "Kepler." To learn more about Johannes Kepler's life and his laws of planetary motion, as well as lots of information on the Kepler Mission, visit NASA's Kepler website and follow the links that interest you.

Key Concepts and Summary

Tycho Brahe's accurate observations of planetary positions provided the data used by Johannes Kepler to derive his three fundamental laws of

planetary motion. Kepler’s laws describe the behavior of planets in their orbits as follows: (1) planetary orbits are ellipses with the Sun at one focus; (2) in equal intervals, a planet’s orbit sweeps out equal areas; and (3) the relationship between the orbital period (P) and the semimajor axis (a) of an orbit is given by $P^2 = a^3$ (when a is in units of AU and P is in units of Earth years).

Glossary

astronomical unit (AU)
: the unit of length defined as the average distance between Earth and the Sun; this distance is about 1.5×10^8 kilometers

eccentricity
: in an ellipse, the ratio of the distance between the foci to the major axis

ellipse
: a closed curve for which the sum of the distances from any point on the ellipse to two points inside (called the foci) is always the same

focus
: (plural: foci) one of two fixed points inside an ellipse from which the sum of the distances to any point on the ellipse is constant

Kepler’s first law
: each planet moves around the Sun in an orbit that is an ellipse, with the Sun at one focus of the ellipse

Kepler’s second law
: the straight line joining a planet and the Sun sweeps out equal areas in space in equal intervals of time

Kepler’s third law
: the square of a planet’s orbital period is directly proportional to the cube of the semimajor axis of its orbit

major axis
: the maximum diameter of an ellipse

orbit
: the path of an object that is in revolution about another object or point

orbital period (*P*)
: the time it takes an object to travel once around the Sun

orbital speed
: the speed at which an object (usually a planet) orbits around the mass of another object; in the case of a planet, the speed at which each planet moves along its ellipse

semimajor axis
: half of the major axis of a conic section, such as an ellipse

Newton’s Great Synthesis

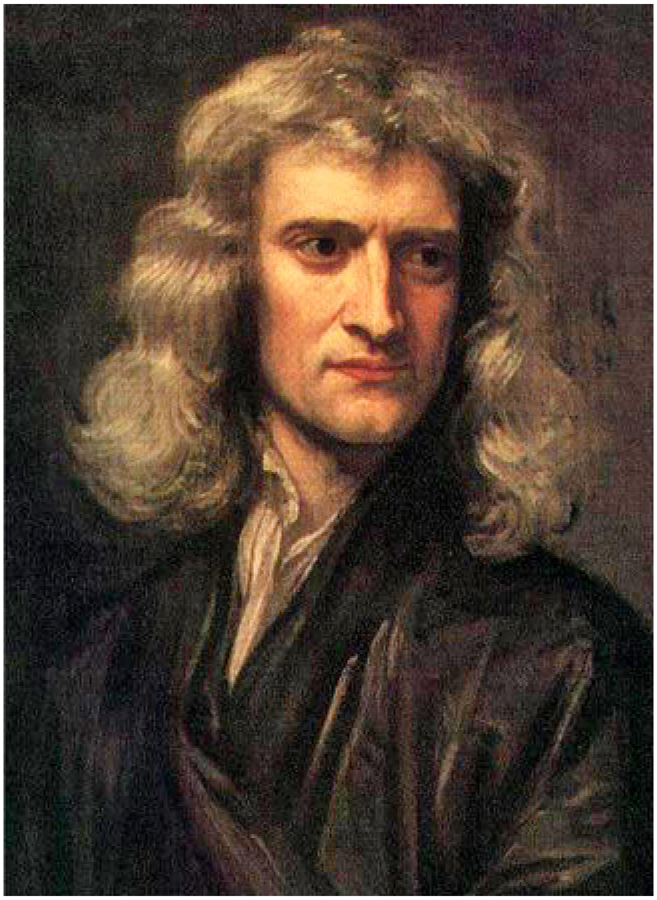
LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe Newton’s three laws of motion
- Explain how Newton’s three laws of motion relate to momentum
- Define mass, volume, and density and how they differ
- Define angular momentum

It was the genius of Isaac Newton that found a conceptual framework that completely explained the observations and rules assembled by Galileo, Brahe, Kepler, and others. Newton was born in Lincolnshire, England, in the year after Galileo’s death ([link]). Against the advice of his mother, who wanted him to stay home and help with the family farm, he entered Trinity College at Cambridge in 1661 and eight years later was appointed professor of mathematics. Among Newton’s contemporaries in England were architect Christopher Wren, authors Aphra Behn and Daniel Defoe, and composer G. F. Handel.

Isaac Newton (1643–1727), 1689 Portrait by Sir Godfrey Kneller.



Isaac Newton's work on the laws of motion, gravity, optics, and mathematics laid the foundations for much of physical science.

Newton's Laws of Motion

As a young man in college, Newton became interested in natural philosophy, as science was then called. He worked out some of his first ideas on machines and optics during the plague years of 1665 and 1666, when students were sent home from college. Newton, a moody and often difficult man, continued to work on his ideas in private, even inventing new mathematical tools to help him deal with the complexities involved. Eventually, his friend Edmund Halley (profiled in Comets and Asteroids:

Debris of the Solar System) prevailed on him to collect and publish the results of his remarkable investigations on motion and gravity. The result was a volume that set out the underlying system of the physical world, *Philosophiae Naturalis Principia Mathematica*. The *Principia*, as the book is generally known, was published at Halley's expense in 1687.

At the very beginning of the *Principia*, Newton proposes three laws that would govern the motions of all objects:

- **Newton's first law**: Every object will continue to be in a state of rest or move at a constant speed in a straight line unless it is compelled to change by an outside force.
- **Newton's second law**: The change of motion of a body is proportional to and in the direction of the force acting on it.
- **Newton's third law**: For every action there is an equal and opposite reaction (*or:* the mutual actions of two bodies upon each other are always equal and act in opposite directions).

In the original Latin, the three laws contain only 59 words, but those few words set the stage for modern science. Let us examine them more carefully.

Interpretation of Newton's Laws

Newton's first law is a restatement of one of Galileo's discoveries, called the *conservation of momentum*. The law states that in the absence of any outside influence, there is a measure of a body's motion, called its **momentum**, that remains unchanged. You may have heard the term momentum used in everyday expressions, such as "This bill in Congress has a lot of momentum; it's going to be hard to stop."

Newton's first law is sometimes called the *law of inertia*, where inertia is the tendency of objects (and legislatures) to keep doing what they are already doing. In other words, a stationary object stays put, and a moving object keeps moving unless some force intervenes.

Let’s define the precise meaning of momentum—it depends on three factors: (1) speed—how fast a body moves (zero if it is stationary), (2) the direction of its motion, and (3) its mass—a measure of the amount of matter in a body, which we will discuss later. Scientists use the term **velocity** to describe the speed and direction of motion. For example, 20 kilometers per hour due south is velocity, whereas 20 kilometers per hour just by itself is speed. Momentum then can be defined as an object’s mass times its velocity.

It’s not so easy to see this rule in action in the everyday world because of the many forces acting on a body at any one time. One important force is friction, which generally slows things down. If you roll a ball along the sidewalk, it eventually comes to a stop because the sidewalk exerts a rubbing force on the ball. But in the space between the stars, where there is so little matter that friction is insignificant, objects can in fact continue to move (to coast) indefinitely.

The momentum of a body can change only under the action of an outside influence. Newton’s second law expresses *force* in terms of its ability to change momentum with time. A force (a push or a pull) has both size and direction. When a force is applied to a body, the momentum changes in the direction of the applied force. This means that a force is required to change either the speed or the direction of a body, or both—that is, to start it moving, to speed it up, to slow it down, to stop it, or to change its direction.

As you learned in Observing the Sky: The Birth of Astronomy, the rate of change in an object’s velocity is called *acceleration*. Newton showed that the acceleration of a body was proportional to the force being applied to it. Suppose that after a long period of reading, you push an astronomy book away from you on a long, smooth table. (We use a smooth table so we can ignore friction.) If you push the book steadily, it will continue to speed up as long as you are pushing it. The harder you push the book, the larger its acceleration will be. How much a force will accelerate an object is also determined by the object’s mass. If you kept pushing a pen with the same force with which you pushed the textbook, the pen—having less mass—would be accelerated to a greater speed.

Newton’s third law is perhaps the most profound of the rules he discovered. Basically, it is a generalization of the first law, but it also gives us a way to define mass. If we consider a system of two or more objects isolated from outside influences, Newton’s first law says that the total momentum of the objects should remain constant. Therefore, any change of momentum within the system must be balanced by another change that is equal and opposite so that the momentum of the entire system is not changed.

This means that forces in nature do not occur alone: we find that in each situation there is always a *pair* of forces that are equal to and opposite each other. If a force is exerted on an object, it must be exerted by something else, and the object will exert an equal and opposite force back on that something. We can look at a simple example to demonstrate this.

Suppose that a daredevil astronomy student—and avid skateboarder—wants to jump from his second-story dorm window onto his board below (we don’t recommend trying this!). The force pulling him down after jumping (as we will see in the next section) is the force of gravity between him and Earth. Both he and Earth must experience the same total change of momentum because of the influence of these mutual forces. So, both the student and Earth are accelerated by each other’s pull. However, the student does much more of the moving. Because Earth has enormously greater mass, it can experience the same change of momentum by accelerating only a very small amount. Things fall toward Earth all the time, but the acceleration of our planet as a result is far too small to be measured.

A more obvious example of the mutual nature of forces between objects is familiar to all who have batted a baseball. The recoil you feel as you swing your bat shows that the ball exerts a force on it during the impact, just as the bat does on the ball. Similarly, when a rifle you are bracing on your shoulder is discharged, the force pushing the bullet out of the muzzle is equal to the force pushing backward upon the gun and your shoulder.

This is the principle behind jet engines and rockets: the force that discharges the exhaust gases from the rear of the rocket is accompanied by the force that pushes the rocket forward. The exhaust gases need not push against air or Earth; a rocket actually operates best in a vacuum ([link]).
Demonstrating Newton’s Third Law.

The U.S. Space Shuttle (here launching *Discovery*), powered by three fuel engines burning liquid oxygen and liquid hydrogen, with two solid fuel boosters, demonstrates Newton’s third law. (credit: modification of work by NASA)

Note:
For more about Isaac Newton’s life and work, check out this timeline page with snapshots from his career, produced by the British Broadcasting Corporation (BBC).

Mass, Volume, and Density

Before we go on to discuss Newton's other work, we want to take a brief look at some terms that will be important to sort out clearly. We begin with *mass*, which is a measure of the amount of material within an object.

The *volume* of an object is the measure of the physical space it occupies. Volume is measured in cubic units, such as cubic centimeters or liters. The volume is the "size" of an object. A penny and an inflated balloon may both have the same mass, but they have very different volumes. The reason is that they also have very different *densities*, which is a measure of how much mass there is per unit volume. Specifically, **density** is the mass divided by the volume. Note that in everyday language we often use "heavy" and "light" as indications of density (rather than weight) as, for instance, when we say that iron is heavy or that whipped cream is light.

The units of density that will be used in this book are grams per cubic centimeter (g/cm^3).[footnote] If a block of some material has a mass of 300 grams and a volume of 100 cm^3, its density is 3 g/cm^3. Familiar materials span a considerable range in density, from artificial materials such as plastic insulating foam (less than 0.1 g/cm^3) to gold (19.3 g/cm^3). [link] gives the densities of some familiar materials. In the astronomical universe, much more remarkable densities can be found, all the way from a comet's tail (10^{-16} g/cm^3) to a collapsed "star corpse" called a neutron star (10^{15} g/cm^3). Generally we use standard metric (or SI) units in this book. The proper metric unit of density in that system is kg/m^3. But to most people, g/cm^3 provides a more meaningful unit because the density of water is exactly 1 g/cm^3, and this is useful information for comparison. Density expressed in g/cm^3 is sometimes called specific density or specific weight.

Densities of Common Materials	
Material	**Density (g/cm^3)**

Densities of Common Materials	
Material	**Density (g/cm^3)**
Gold	19.3
Lead	11.3
Iron	7.9
Earth (bulk)	5.5
Rock (typical)	2.5
Water	1
Wood (typical)	0.8
Insulating foam	0.1
Silica gel	0.02

To sum up, mass is *how much*, volume is *how big*, and density is *how tightly packed*.

Note:
You can play with a simple animation demonstrating the relationship between the concepts of density, mass, and volume, and find out why objects like wood float in water.

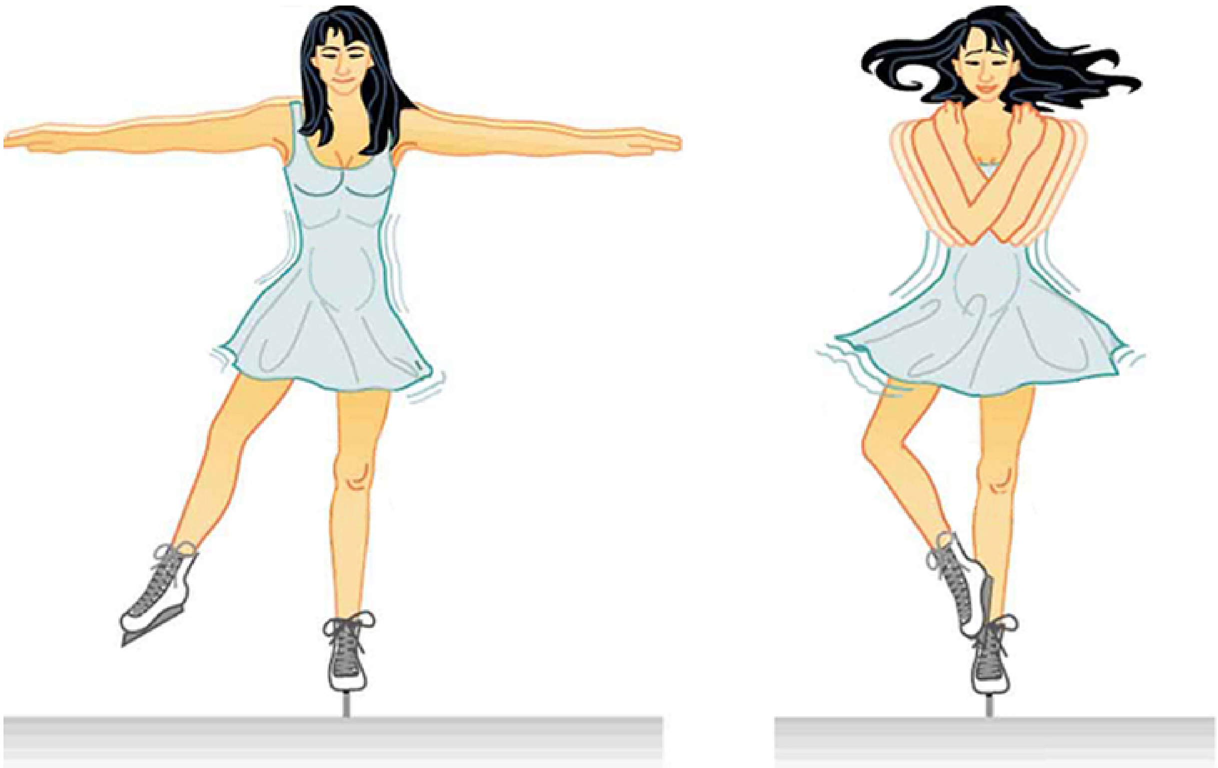
Angular Momentum

A concept that is a bit more complex, but important for understanding many astronomical objects, is **angular momentum**, which is a measure of the rotation of a body as it revolves around some fixed point (an example is a planet orbiting the Sun). The angular momentum of an object is defined as the product of its mass, its velocity, and its distance from the fixed point around which it revolves.

If these three quantities remain constant—that is, if the motion of a particular object takes place at a constant velocity at a fixed distance from the spin center—then the angular momentum is also a constant. Kepler’s second law is a consequence of the *conservation of angular momentum*. As a planet approaches the Sun on its elliptical orbit and the distance to the spin center decreases, the planet speeds up to conserve the angular momentum. Similarly, when the planet is farther from the Sun, it moves more slowly.

The conservation of angular momentum is illustrated by figure skaters, who bring their arms and legs in to spin more rapidly, and extend their arms and legs to slow down ([link]). You can duplicate this yourself on a well-oiled swivel stool by starting yourself spinning slowly with your arms extended and then pulling your arms in. Another example of the conservation of angular momentum is a shrinking cloud of dust or a star collapsing on itself (both are situations that you will learn about as you read on). As material moves to a lesser distance from the spin center, the speed of the material increases to conserve angular momentum.

Conservation of Angular Momentum.



When a spinning figure skater brings in her arms, their distance from her spin center is smaller, so her speed increases. When her arms are out, their distance from the spin center is greater, so she slows down.

Key Concepts and Summary

In his *Principia*, Isaac Newton established the three laws that govern the motion of objects: (1) objects continue to be at rest or move with a constant velocity unless acted upon by an outside force; (2) an outside force causes an acceleration (and changes the momentum) for an object; and (3) for every action there is an equal and opposite reaction. Momentum is a measure of the motion of an object and depends on both its mass and its velocity. Angular momentum is a measure of the motion of a spinning or revolving object and depends on its mass, velocity, and distance from the point around which it revolves. The density of an object is its mass divided by its volume.

Glossary

angular momentum
: the measure of the motion of a rotating object in terms of its speed and how widely the object’s mass is distributed around its axis

density
: the ratio of the mass of an object to its volume

momentum
: the measure of the amount of motion of a body; the momentum of a body is the product of its mass and velocity; in the absence of an unbalanced force, momentum is conserved

Newton’s first law
: every object will continue to be in a state of rest or move at a constant speed in a straight line unless it is compelled to change by an outside force

Newton’s second law
: the change of motion of a body is proportional to and in the direction of the force acting on it

Newton’s third law
: for every action there is an equal and opposite reaction (*or:* the mutual actions of two bodies upon each other are always equal and act in opposite directions)

velocity
: the speed and direction a body is moving—for example, 44 kilometers per second toward the north galactic pole

Newton’s Universal Law of Gravitation

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain what determines the strength of gravity
- Describe how Newton’s universal law of gravitation extends our understanding of Kepler’s laws

Newton’s laws of motion show that objects at rest will stay at rest and those in motion will continue moving uniformly in a straight line unless acted upon by a force. Thus, it is the *straight line* that defines the most natural state of motion. But the planets move in ellipses, not straight lines; therefore, some force must be bending their paths. That force, Newton proposed, was **gravity**.

In Newton’s time, gravity was something associated with Earth alone. Everyday experience shows us that Earth exerts a gravitational force upon objects at its surface. If you drop something, it accelerates toward Earth as it falls. Newton’s insight was that Earth’s gravity might extend as far as the Moon and produce the force required to curve the Moon’s path from a straight line and keep it in its orbit. He further hypothesized that gravity is not limited to Earth, but that there is a general force of attraction between all material bodies. If so, the attractive force between the Sun and each of the planets could keep them in their orbits. (This may seem part of our everyday thinking today, but it was a remarkable insight in Newton’s time.)

Once Newton boldly hypothesized that there was a universal attraction among all bodies everywhere in space, he had to determine the exact nature of the attraction. The precise mathematical description of that gravitational force had to dictate that the planets move exactly as Kepler had described them to (as expressed in Kepler's three laws). Also, that gravitational force had to predict the correct behavior of falling bodies on Earth, as observed by Galileo. How must the force of gravity depend on distance in order for these conditions to be met?

The answer to this question required mathematical tools that had not yet been developed, but this did not deter Isaac Newton, who invented what we today call calculus to deal with this problem. Eventually he was able to conclude that the magnitude of the force of gravity must decrease with increasing distance between the Sun and a planet (or between any two objects) in proportion to the inverse square of their separation. In other words, if a planet were twice as far from the Sun, the force would be $(1/2)^2$, or 1/4 as large. Put the planet three times farther away, and the force is $(1/3)^2$, or 1/9 as large.

Newton also concluded that the gravitational attraction between two bodies must be proportional to their masses. The more mass an object has, the stronger the pull of its gravitational force. The gravitational attraction between any two objects is therefore given by one of the most famous equations in all of science:

Equation:

$$F_{\text{gravity}} = G\frac{M_1 M_2}{R^2}$$

where F_{gravity} is the gravitational force between two objects, M_1 and M_2 are the masses of the two objects, and R is their separation. G is a constant number known as the *universal gravitational constant*, and the equation itself symbolically summarizes Newton's *universal law of gravitation*. With such a force and the laws of motion, Newton was able to show mathematically that the only orbits permitted were exactly those described by Kepler's laws.

Newton’s universal law of gravitation works for the planets, but is it really universal? The gravitational theory should also predict the observed acceleration of the Moon toward Earth as it orbits Earth, as well as of any object (say, an apple) dropped near Earth’s surface. The falling of an apple is something we can measure quite easily, but can we use it to predict the motions of the Moon?

Recall that according to Newton’s second law, forces cause acceleration. Newton’s universal law of gravitation says that the force acting upon (and therefore the acceleration of) an object toward Earth should be inversely proportional to the square of its distance from the center of Earth. Objects like apples at the surface of Earth, at a distance of one Earth-radius from the center of Earth, are observed to accelerate downward at 9.8 meters per second per second (9.8 m/s^2).

It is this force of gravity on the surface of Earth that gives us our sense of *weight*. Unlike your mass, which would remain the same on any planet or moon, your weight depends on the local force of gravity. So you would weigh less on Mars and the Moon than on Earth, even though there is no change in your mass. (Which means you would still have to go easy on the desserts in the college cafeteria when you got back!)

The Moon is 60 Earth radii away from the center of Earth. If gravity (and the acceleration it causes) gets weaker with distance squared, the acceleration the Moon experiences should be a lot less than for the apple. The acceleration should be $(1/60)^2 = 1/3600$ (or 3600 times less—about 0.00272 m/s^2. This is precisely the observed acceleration of the Moon in its orbit. (As we shall see, the Moon does not fall *to* Earth with this acceleration, but falls *around* Earth.) Imagine the thrill Newton must have felt to realize he had discovered, and verified, a law that holds for Earth, apples, the Moon, and, as far as he knew, everything in the universe.

Example:

Calculating Weight

By what factor would a person’s weight at the surface of Earth change if Earth had its present mass but eight times its present volume?

Solution

With eight times the volume, Earth’s radius would double. This means the gravitational force at the surface would reduce by a factor of $(1/2)^2 = 1/4$, so a person would weigh only one-fourth as much.

Check Your Learning

By what factor would a person’s weight at the surface of Earth change if Earth had its present size but only one-third its present mass?

Note:

Answer:

With one-third its present mass, the gravitational force at the surface would reduce by a factor of 1/3, so a person would weight only one-third as much.

Gravity is a “built-in” property of mass. Whenever there are masses in the universe, they will interact via the force of gravitational attraction. The more mass there is, the greater the force of attraction. Here on Earth, the largest concentration of mass is, of course, the planet we stand on, and its pull dominates the gravitational interactions we experience. But everything with mass attracts everything else with mass anywhere in the universe.

Newton’s law also implies that gravity never becomes zero. It quickly gets weaker with distance, but it continues to act to some degree no matter how far away you get. The pull of the Sun is stronger at Mercury than at Pluto, but it can be felt far beyond Pluto, where astronomers have good evidence that it continuously makes enormous numbers of smaller icy bodies move around huge orbits. And the Sun’s gravitational pull joins with the pull of billions of others stars to create the gravitational pull of our Milky Way Galaxy. That force, in turn, can make other smaller galaxies orbit around the Milky Way, and so on.

Why is it then, you may ask, that the astronauts aboard the Space Shuttle appear to have no gravitational forces acting on them when we see images on television of the astronauts and objects floating in the spacecraft? After all, the astronauts in the shuttle are only a few hundred kilometers above the surface of Earth, which is not a significant distance compared to the size of Earth, so gravity is certainly not a great deal weaker that much farther away. The astronauts feel “weightless” (meaning that they don’t feel the gravitational force acting on them) for the same reason that passengers in an elevator whose cable has broken or in an airplane whose engines no longer work feel weightless: they are falling ([link]).[footnote]

In the film *Apollo 13*, the scenes in which the astronauts were “weightless” were actually filmed in a falling airplane. As you might imagine, the plane fell for only short periods before the engines engaged again.

Astronauts in Free Fall.

While in space, astronauts are falling freely, so they experience “weightlessness.” Clockwise from top left: Tracy Caldwell Dyson (NASA), Naoko Yamzaki (JAXA), Dorothy Metcalf-Lindenburger (NASA), and Stephanie Wilson (NASA). (credit: NASA)

When *falling*, they are in free fall and accelerate at the same rate as everything around them, including their spacecraft or a camera with which they are taking photographs of Earth. When doing so, astronauts experience no additional forces and therefore feel “weightless.” Unlike the falling elevator passengers, however, the astronauts are falling *around* Earth, not *to* Earth; as a result they will continue to fall and are said to be “in orbit” around Earth (see the next section for more about orbits).

Orbital Motion and Mass

Kepler’s laws describe the orbits of the objects whose motions are described by Newton’s laws of motion and the law of gravity. Knowing that gravity is the force that attracts planets toward the Sun, however, allowed Newton to rethink Kepler’s third law. Recall that Kepler had found a relationship between the orbital period of a planet’s revolution and its distance from the Sun. But Newton’s formulation introduces the additional factor of the masses of the Sun (M_1) and the planet (M_2), both expressed in units of the Sun’s mass. Newton’s universal law of gravitation can be used to show mathematically that this relationship is actually
Equation:

$$a^3 = \left(M_1 + M_2\right) \times P^2$$

where *a* is the semimajor axis and *P* is the orbital period.

How did Kepler miss this factor? In units of the Sun’s mass, the mass of the Sun is 1, and in units of the Sun’s mass, the mass of a typical planet is a negligibly small factor. This means that the sum of the Sun’s mass and a planet’s mass, (M_1 + M_2), is very, very close to 1. This makes Newton’s formula appear almost the same as Kepler’s; the tiny mass of the planets compared to the Sun is the reason that Kepler did not realize that both masses had to be included in the calculation. There are many situations in astronomy, however, in which we *do* need to include the two mass terms—for example, when two stars or two galaxies orbit each other.

Including the mass term allows us to use this formula in a new way. If we can measure the motions (distances and orbital periods) of objects acting under their mutual gravity, then the formula will permit us to deduce their masses. For example, we can calculate the mass of the Sun by using the distances and orbital periods of the planets, or the mass of Jupiter by noting the motions of its moons.

Indeed, Newton's reformulation of Kepler's third law is one of the most powerful concepts in astronomy. Our ability to deduce the masses of objects from their motions is key to understanding the nature and evolution of many astronomical bodies. We will use this law repeatedly throughout this text in calculations that range from the orbits of comets to the interactions of galaxies.

Example:

Calculating the Effects of Gravity

A planet like Earth is found orbiting its star at a distance of 1 AU in 0.71 Earth-year. Can you use Newton's version of Kepler's third law to find the mass of the star? (Remember that compared to the mass of a star, the mass of an earthlike planet can be considered negligible.)

Solution

In the formula $a^3 = (M_1 + M_2) \times P^2$, the factor $M_1 + M_2$ would now be approximately equal to M_1 (the mass of the star), since the planet's mass is so small by comparison. Then the formula becomes $a^3 = M_1 \times P^2$, and we can solve for M_1:

Equation:

$$M_1 = \frac{a^3}{P^2}$$

Since $a = 1$, $a^3 = 1$, so

Equation:

$$M_1 = \frac{1}{P^2} = \frac{1}{0.71^2} = \frac{1}{0.5} = 2$$

So the mass of the star is twice the mass of our Sun. (Remember that this way of expressing the law has units in terms of Earth and the Sun, so masses are expressed in units of the mass of our Sun.)

Check Your Learning

Suppose a star with twice the mass of our Sun had an earthlike planet that took 4 years to orbit the star. At what distance (semimajor axis) would this planet orbit its star?

Note:

Answer:

Again, we can neglect the mass of the planet. So $M_1 = 2$ and $P = 4$ years. The formula is $a^3 = M_1 \times P^2$, so $a^3 = 2 \times 4^2 = 2 \times 16 = 32$. So a is the cube root of 32. To find this, you can just ask Google, “What is the cube root of 32?” and get the answer 3.2 AU.

Note:

You might like to try a simulation that lets you move the Sun, Earth, Moon, and space station to see the effects of changing their distances on their gravitational forces and orbital paths. You can even turn off gravity and see what happens.

Key Concepts and Summary

Gravity, the attractive force between all masses, is what keeps the planets in orbit. Newton’s universal law of gravitation relates the gravitational force to mass and distance:

Equation:

$$F_{\text{gravity}} = G\frac{M_1 M_2}{R^2}$$

The force of gravity is what gives us our sense of weight. Unlike mass, which is constant, weight can vary depending on the force of gravity (or acceleration) you feel. When Kepler's laws are reexamined in the light of Newton's gravitational law, it becomes clear that the masses of both objects are important for the third law, which becomes $a^3 = (M_1 + M_2) \times P^2$. Mutual gravitational effects permit us to calculate the masses of astronomical objects, from comets to galaxies.

Glossary

gravity
: the mutual attraction of material bodies or particles

Orbits in the Solar System

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Compare the orbital characteristics of the planets in the solar system
- Compare the orbital characteristics of asteroids and comets in the solar system

Recall that the path of an object under the influence of gravity through space is called its orbit, whether that object is a spacecraft, planet, star, or galaxy. An orbit, once determined, allows the future positions of the object to be calculated.

Two points in any orbit in our solar system have been given special names. The place where the planet is closest to the Sun (*helios* in Greek) and moves the fastest is called the **perihelion** of its orbit, and the place where it is farthest away and moves the most slowly is the **aphelion**. For the Moon or a satellite orbiting Earth (*gee* in Greek), the corresponding terms are **perigee** and **apogee**. (In this book, we use the word *moon* for a natural object that goes around a planet and the word **satellite** to mean a human-made object that revolves around a planet.)

Orbits of the Planets

Today, Newton's work enables us to calculate and predict the orbits of the planets with marvelous precision. We know eight planets, beginning with

Mercury closest to the Sun and extending outward to Neptune. The average orbital data for the planets are summarized in [link]. (Ceres is the largest of the *asteroids*, now considered a dwarf planet.)

According to Kepler's laws, Mercury must have the shortest orbital period (88 Earth-days); thus, it has the highest orbital speed, averaging 48 kilometers per second. At the opposite extreme, Neptune has a period of 165 years and an average orbital speed of just 5 kilometers per second.

All the planets have orbits of rather low eccentricity. The most eccentric orbit is that of Mercury (0.21); the rest have eccentricities smaller than 0.1. It is fortunate that among the rest, Mars has an eccentricity greater than that of many of the other planets. Otherwise the pre-telescopic observations of Brahe would not have been sufficient for Kepler to deduce that its orbit had the shape of an ellipse rather than a circle.

The planetary orbits are also confined close to a common plane, which is near the plane of Earth's orbit (called the ecliptic). The strange orbit of the dwarf planet Pluto is inclined about 17° to the ecliptic, and that of the dwarf planet Eris (orbiting even farther away from the Sun than Pluto) by 44°, but all the major planets lie within 10° of the common plane of the solar system.

Note:
You can use an orbital simulator to design your own mini solar system with up to four bodies. Adjust masses, velocities, and positions of the planets, and see what happens to their orbits as a result.

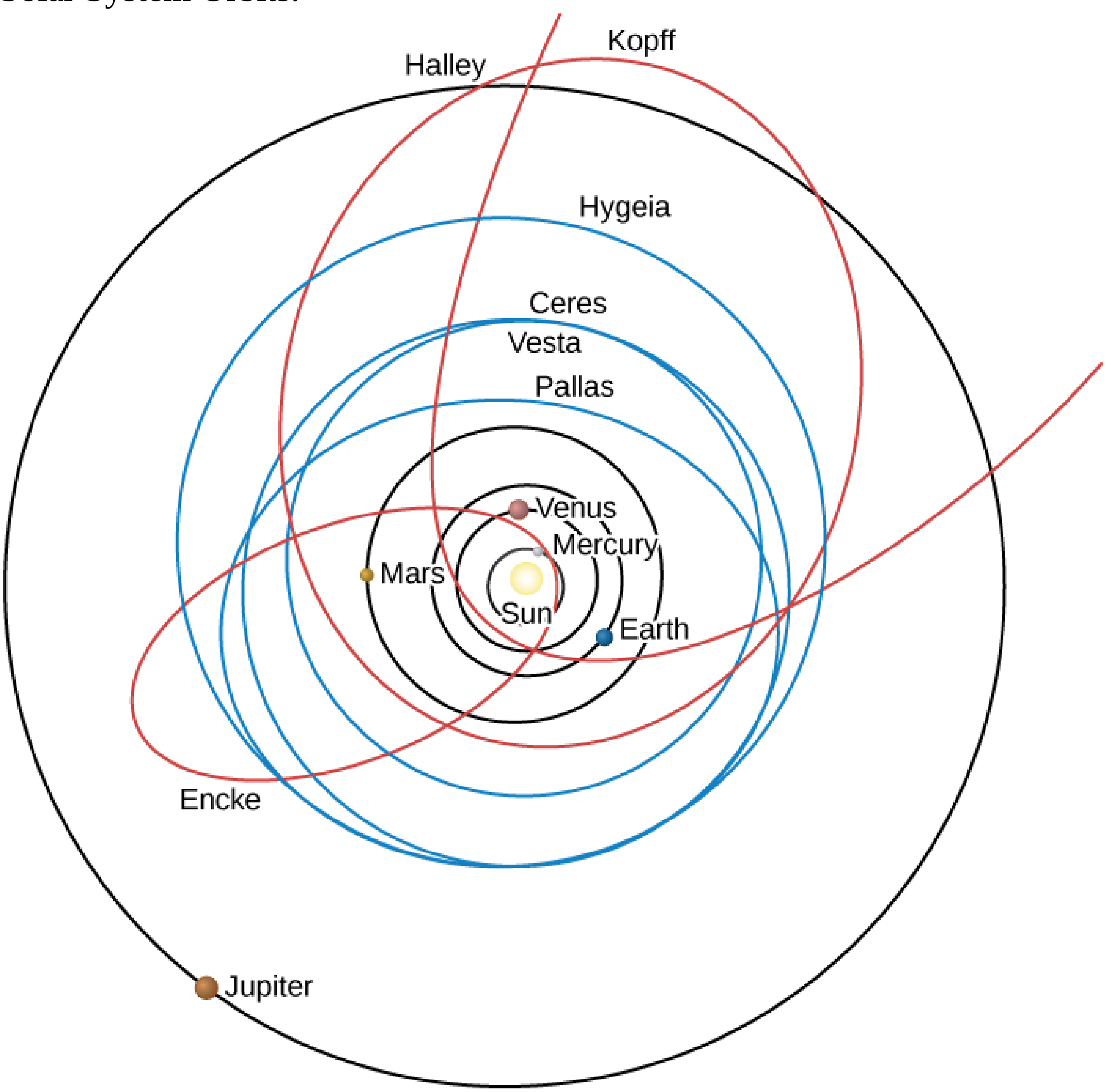
Orbits of Asteroids and Comets

In addition to the eight planets, there are many smaller objects in the solar system. Some of these are moons (natural satellites) that orbit all the planets except Mercury and Venus. In addition, there are two classes of smaller objects in heliocentric orbits: *asteroids* and *comets*. Both asteroids and

comets are believed to be small chunks of material left over from the formation process of the solar system.

In general, asteroids have orbits with smaller semimajor axes than do comets ([link]). The majority of them lie between 2.2 and 3.3 AU, in the region known as the **asteroid belt** (see Comets and Asteroids: Debris of the Solar System). As you can see in [link], the asteroid belt (represented by its largest member, Ceres) is in the middle of a gap between the orbits of Mars and Jupiter. It is because these two planets are so far apart that stable orbits of small bodies can exist in the region between them.

Solar System Orbits.



We see the orbits of typical comets and asteroids compared with those of the planets Mercury, Venus, Earth, Mars, and Jupiter (black circles). Shown in red are three comets: Halley, Kopff, and Encke. In blue are the four largest asteroids: Ceres, Pallas, Vesta, and Hygeia.

Orbital Data for the Planets			
Planet	**Semimajor Axis (AU)**	**Period (y)**	**Eccentricity**
Mercury	0.39	0.24	0.21
Venus	0.72	0.6	0.01
Earth	1	1.00	0.02
Mars	1.52	1.88	0.09
(Ceres)	2.77	4.6	0.08
Jupiter	5.20	11.86	0.05
Saturn	9.54	29.46	0.06
Uranus	19.19	84.01	0.05
Neptune	30.06	164.82	0.01

Comets generally have orbits of larger size and greater eccentricity than those of the asteroids. Typically, the eccentricity of their orbits is 0.8 or higher. According to Kepler's second law, therefore, they spend most of their time far from the Sun, moving very slowly. As they approach perihelion, the comets speed up and whip through the inner parts of their orbits more rapidly.

Key Concepts and Summary

The closest point in a satellite orbit around Earth is its perigee, and the farthest point is its apogee (corresponding to perihelion and aphelion for an orbit around the Sun). The planets follow orbits around the Sun that are nearly circular and in the same plane. Most asteroids are found between Mars and Jupiter in the asteroid belt, whereas comets generally follow orbits of high eccentricity.

Glossary

aphelion
: the point in its orbit where a planet (or other orbiting object) is farthest from the Sun

apogee
: the point in its orbit where an Earth satellite is farthest from Earth

asteroid belt
: the region of the solar system between the orbits of Mars and Jupiter in which most asteroids are located; the main belt, where the orbits are generally the most stable, extends from 2.2 to 3.3 AU from the Sun

perigee
: the point in its orbit where an Earth satellite is closest to Earth

perihelion
: the point in its orbit where a planet (or other orbiting object) is nearest to the Sun

satellite

an object that revolves around a planet

Motions of Satellites and Spacecraft

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how an object (such as a satellite) can be put into orbit around Earth
- Explain how an object (such as a planetary probe) can escape from orbit

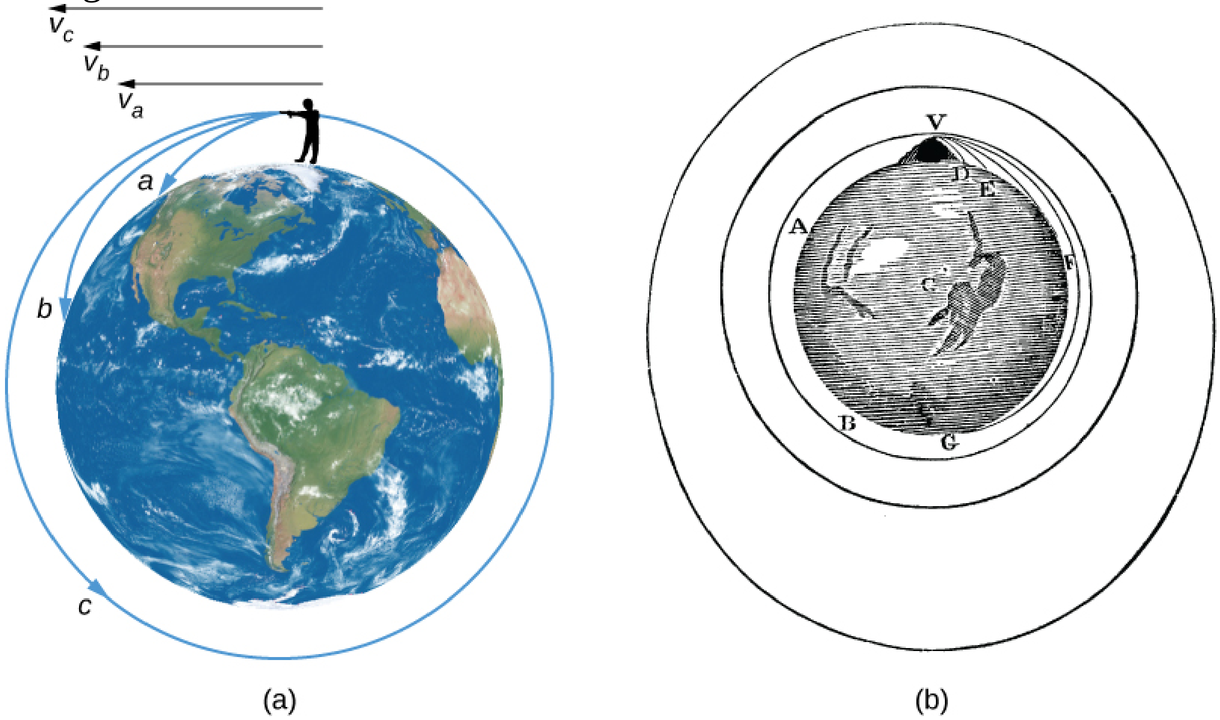
Newton's universal law of gravitation and Kepler's laws describe the motions of Earth satellites and interplanetary spacecraft as well as the planets. Sputnik, the first artificial Earth satellite, was launched by what was then called the Soviet Union on October 4, 1957. Since that time, thousands of satellites have been placed into orbit around Earth, and spacecraft have also orbited the Moon, Venus, Mars, Jupiter, Saturn, and a number of asteroids and comets.

Once an artificial satellite is in orbit, its behavior is no different from that of a natural satellite, such as our Moon. If the satellite is high enough to be free of atmospheric friction, it will remain in orbit forever. However, although there is no difficulty in maintaining a satellite once it is in orbit, a great deal of energy is required to lift the spacecraft off Earth and accelerate it to orbital speed.

To illustrate how a satellite is launched, imagine a gun firing a bullet horizontally from the top of a high mountain, as in [link], which has been adapted from a similar diagram by Newton. Imagine, further, that the

friction of the air could be removed and that nothing gets in the bullet's way. Then the only force that acts on the bullet after it leaves the muzzle is the gravitational force between the bullet and Earth.

Firing a Bullet into Orbit.



(a) For paths *a* and *b*, the velocity is not enough to prevent gravity from pulling the bullet back to Earth; in case *c*, the velocity allows the bullet to fall completely around Earth. (b) This diagram by Newton in his *De Mundi Systemate*, 1731 edition, illustrates the same concept shown in (a).

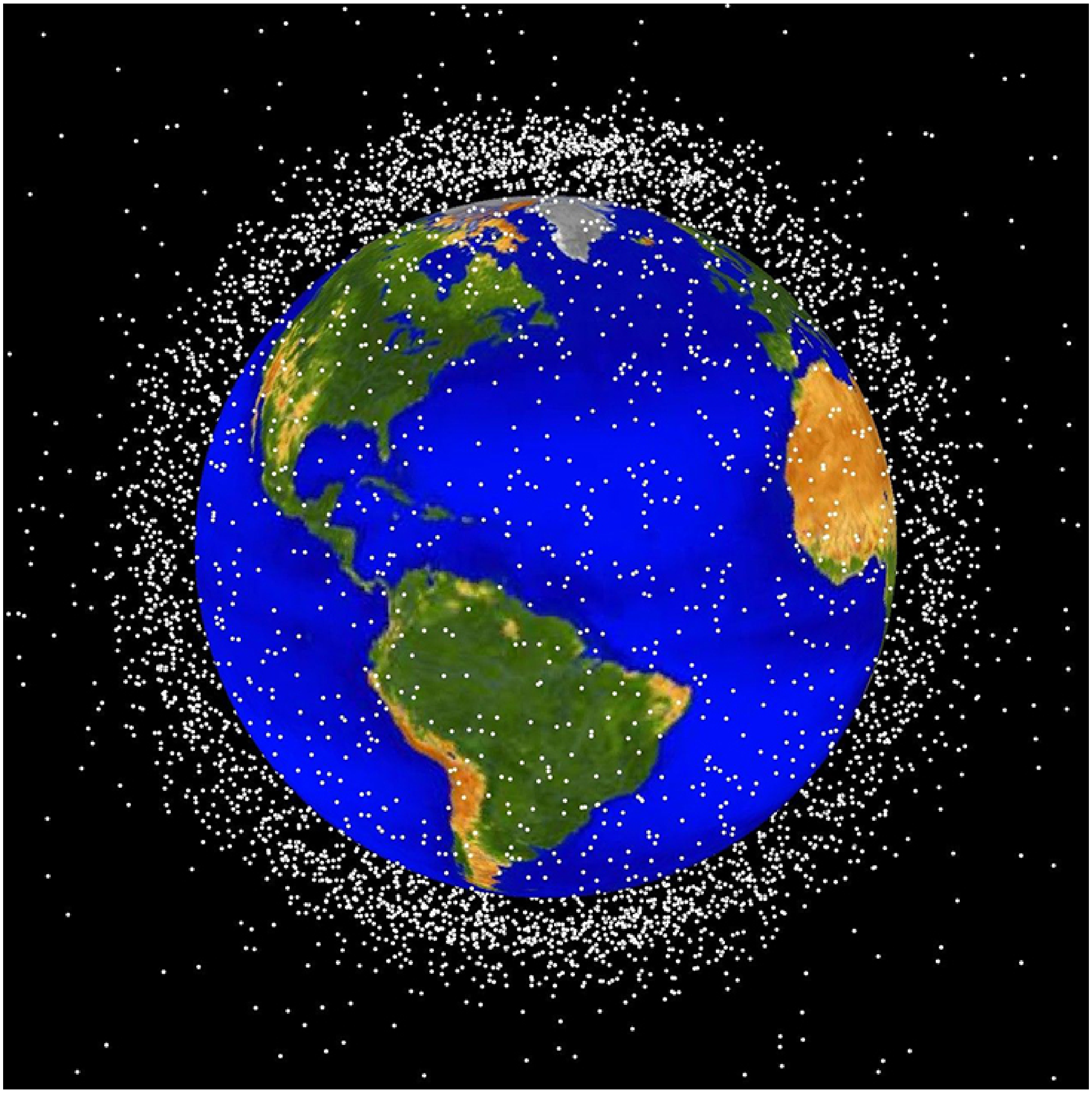
If the bullet is fired with a velocity we can call v_a, the gravitational force acting upon it pulls it downward toward Earth, where it strikes the ground at point *a*. However, if it is given a higher muzzle velocity, v_b, its higher speed carries it farther before it hits the ground at point *b*.

If our bullet is given a high enough muzzle velocity, v_c, the curved surface of Earth causes the ground to remain the same distance from the bullet so that the bullet falls *around* Earth in a complete circle. The speed needed to

do this—called the circular satellite velocity—is about 8 kilometers per second, or about 17,500 miles per hour in more familiar units.

Each year, more than 50 new satellites are launched into orbit by such nations as Russia, the United States, China, Japan, India, and Israel, as well as by the European Space Agency (ESA), a consortium of European nations ([link]). Today, these satellites are used for weather tracking, ecology, global positioning systems, communications, and military purposes, to name a few uses. Most satellites are launched into low Earth orbit, since this requires the minimum launch energy. At the orbital speed of 8 kilometers per second, they circle the planet in about 90 minutes. Some of the very low Earth orbits are not indefinitely stable because, as Earth's atmosphere swells from time to time, a frictional drag is generated by the atmosphere on these satellites, eventually leading to a loss of energy and "decay" of the orbit.

Satellites in Earth Orbit.



This figure shows the larger pieces of orbital debris that are being tracked by NASA in Earth’s orbit. (credit: NASA/JSC)

Interplanetary Spacecraft

The exploration of the solar system has been carried out largely by robot spacecraft sent to the other planets. To escape Earth, these craft must

achieve **escape speed**, the speed needed to move away from Earth forever, which is about 11 kilometers per second (about 25,000 miles per hour). After escaping Earth, these craft coast to their targets, subject only to minor trajectory adjustments provided by small thruster rockets on board. In interplanetary flight, these spacecraft follow orbits around the Sun that are modified only when they pass near one of the planets.

As it comes close to its target, a spacecraft is deflected by the planet's gravitational force into a modified orbit, either gaining or losing energy in the process. Spacecraft controllers have actually been able to use a planet's gravity to redirect a flyby spacecraft to a second target. For example, Voyager 2 used a series of gravity-assisted encounters to yield successive flybys of Jupiter (1979), Saturn (1980), Uranus (1986), and Neptune (1989). The Galileo spacecraft, launched in 1989, flew past Venus once and Earth twice to gain the energy required to reach its ultimate goal of orbiting Jupiter.

If we wish to orbit a planet, we must slow the spacecraft with a rocket when the spacecraft is near its destination, allowing it to be captured into an elliptical orbit. Additional rocket thrust is required to bring a vehicle down from orbit for a landing on the surface. Finally, if a return trip to Earth is planned, the landed payload must include enough propulsive power to repeat the entire process in reverse.

Key Concepts and Summary

The orbit of an artificial satellite depends on the circumstances of its launch. The circular satellite velocity needed to orbit Earth's surface is 8 kilometers per second, and the escape speed from our planet is 11 kilometers per second. There are many possible interplanetary trajectories, including those that use gravity-assisted flybys of one object to redirect the spacecraft toward its next target.

Glossary

escape speed

the speed a body must achieve to break away from the gravity of another body

Gravity with More Than Two Bodies

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how the gravitational interactions of many bodies can causes perturbations in their motions
- Explain how the planet Neptune was discovered

Until now, we have considered the Sun and a planet (or a planet and one of its moons) as nothing more than a pair of bodies revolving around each other. In fact, all the planets exert gravitational forces upon one another as well. These interplanetary attractions cause slight variations from the orbits than would be expected if the gravitational forces between planets were neglected. The motion of a body that is under the gravitational influence of two or more other bodies is very complicated and can be calculated properly only with large computers. Fortunately, astronomers have such computers at their disposal in universities and government research institutes.

The Interactions of Many Bodies

As an example, suppose you have a cluster of a thousand stars all orbiting a common center (such clusters are quite common, as we shall see in Star Clusters). If we know the exact position of each star at any given instant, we can calculate the combined gravitational force of the entire group on any one member of the cluster. Knowing the force on the star in question, we can therefore find how it will accelerate. If we know how it was moving to

begin with, we can then calculate how it will move in the next instant of time, thus tracking its motion.

However, the problem is complicated by the fact that the other stars are also moving and thus changing the effect they will have on our star. Therefore, we must simultaneously calculate the acceleration of each star produced by the combination of the gravitational attractions of all the others in order to track the motions of all of them, and hence of any one. Such complex calculations have been carried out with modern computers to track the evolution of hypothetical clusters of stars with up to a million members ([link]).

Modern Computing Power.

These supercomputers at NASA's Ames Research Center are capable of tracking the motions of more than a million objects under their mutual gravitation. (credit: NASA Ames Research Center/Tom Trower)

Within the solar system, the problem of computing the orbits of planets and spacecraft is somewhat simpler. We have seen that Kepler's laws, which do

not take into account the gravitational effects of the other planets on an orbit, really work quite well. This is because these additional influences are very small in comparison with the dominant gravitational attraction of the Sun. Under such circumstances, it is possible to treat the effects of other bodies as small **perturbations** (or disturbances). During the eighteenth and nineteenth centuries, mathematicians developed many elegant techniques for calculating perturbations, permitting them to predict very precisely the positions of the planets. Such calculations eventually led to the prediction and discovery of a new planet in 1846.

The Discovery of Neptune

The discovery of the eighth planet, Neptune, was one of the high points in the development of gravitational theory. In 1781, William Herschel, a musician and amateur astronomer, accidentally discovered the seventh planet, Uranus. It happens that Uranus had been observed a century before, but in none of those earlier sightings was it recognized as a planet; rather, it was simply recorded as a star. Herschel’s discovery showed that there could be planets in the solar system too dim to be visible to the unaided eye, but ready to be discovered with a telescope if we just knew where to look.

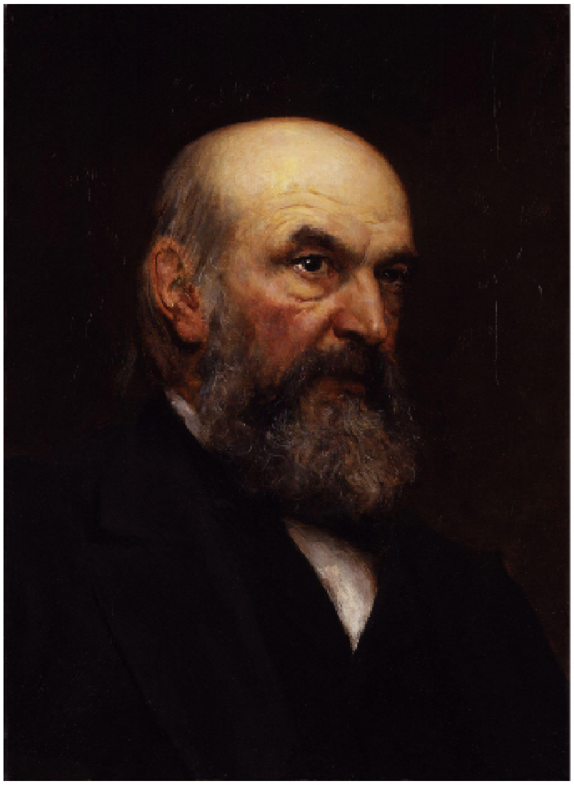
By 1790, an orbit had been calculated for Uranus using observations of its motion in the decade following its discovery. Even after allowance was made for the perturbing effects of Jupiter and Saturn, however, it was found that Uranus did not move on an orbit that exactly fit the earlier observations of it made since 1690. By 1840, the discrepancy between the positions observed for Uranus and those predicted from its computed orbit amounted to about 0.03°—an angle barely discernable to the unaided eye but still larger than the probable errors in the orbital calculations. In other words, Uranus just did not seem to move on the orbit predicted from Newtonian theory.

In 1843, John Couch Adams, a young Englishman who had just completed his studies at Cambridge, began a detailed mathematical analysis of the irregularities in the motion of Uranus to see whether they might be produced by the pull of an unknown planet. He hypothesized a planet more distant from the Sun than Uranus, and then determined the mass and orbit it

had to have to account for the departures in Uranus' orbit. In October 1845, Adams delivered his results to George Airy, the British Astronomer Royal, informing him where in the sky to find the new planet. We now know that Adams' predicted position for the new body was correct to within 2°, but for a variety of reasons, Airy did not follow up right away.

Meanwhile, French mathematician Urbain Jean Joseph Le Verrier, unaware of Adams or his work, attacked the same problem and published its solution in June 1846. Airy, noting that Le Verrier's predicted position for the unknown planet agreed to within 1° with that of Adams, suggested to James Challis, Director of the Cambridge Observatory, that he begin a search for the new object. The Cambridge astronomer, having no up-to-date star charts of the Aquarius region of the sky where the planet was predicted to be, proceeded by recording the positions of all the faint stars he could observe with his telescope in that location. It was Challis' plan to repeat such plots at intervals of several days, in the hope that the planet would distinguish itself from a star by its motion. Unfortunately, he was negligent in examining his observations; although he had actually seen the planet, he did not recognize it.

About a month later, Le Verrier suggested to Johann Galle, an astronomer at the Berlin Observatory, that he look for the planet. Galle received Le Verrier's letter on September 23, 1846, and, possessing new charts of the Aquarius region, found and identified the planet that very night. It was less than a degree from the position Le Verrier predicted. The discovery of the eighth planet, now known as Neptune (the Latin name for the god of the sea), was a major triumph for gravitational theory for it dramatically confirmed the generality of Newton's laws. The honor for the discovery is properly shared by the two mathematicians, Adams and Le Verrier ([link]).
Mathematicians Who Discovered a Planet.



(a)

(b)

(a) John Couch Adams (1819–1892) and (b) Urbain J. J. Le Verrier (1811–1877) share the credit for discovering the planet Neptune.

We should note that the discovery of Neptune was not a complete surprise to astronomers, who had long suspected the existence of the planet based on the “disobedient” motion of Uranus. On September 10, 1846, two weeks before Neptune was actually found, John Herschel, son of the discoverer of Uranus, remarked in a speech before the British Association, “We see [the new planet] as Columbus saw America from the shores of Spain. Its movements have been felt trembling along the far-reaching line of our analysis with a certainty hardly inferior to ocular demonstration.”

This discovery was a major step forward in combining Newtonian theory with painstaking observations. Such work continues in our own times with the discovery of planets around other stars.

Note:
For the fuller story of how Neptune was predicted and found (and the effect of the discovery on the search for Pluto), you can read this page on the mathematical discovery of planets.

Note:
Astronomy and the Poets
When Copernicus, Kepler, Galileo, and Newton formulated the fundamental rules that underlie everything in the physical world, they changed much more than the face of science. For some, they gave humanity the courage to let go of old superstitions and see the world as rational and manageable; for others, they upset comforting, ordered ways that had served humanity for centuries, leaving only a dry, “mechanical clockwork” universe in their wake.
Poets of the time reacted to such changes in their work and debated whether the new world picture was an appealing or frightening one. John Donne (1573–1631), in a poem called “Anatomy of the World,” laments the passing of the old certainties:

> The new Philosophy [science] calls all in doubt,
> The element of fire is quite put out;
> The Sun is lost, and th’ earth, and no man’s wit
> Can well direct him where to look for it.

(Here the “element of fire” refers also to the sphere of fire, which medieval thought placed between Earth and the Moon.)
By the next century, however, poets like Alexander Pope were celebrating Newton and the Newtonian world view. Pope’s famous couplet, written upon Newton’s death, goes

> Nature, and nature’s laws lay hid in night.
> God said, Let Newton be! And all was light.

In his 1733 poem, *An Essay on Man*, Pope delights in the complexity of the new views of the world, incomplete though they are:

> Of man, what see we, but his station here,
> From which to reason, to which refer? . . .
> He, who thro’ vast immensity can pierce,
> See worlds on worlds compose one universe,
> Observe how system into system runs,
> What other planets circle other suns,
> What vary’d being peoples every star,
> May tell why Heav’n has made us as we are . . .
> All nature is but art, unknown to thee;
> All chance, direction, which thou canst not see;
> All discord, harmony not understood;
> All partial evil, universal good:
> And, in spite of pride, in erring reason’s spite,
> One truth is clear, whatever is, is right.

Poets and philosophers continued to debate whether humanity was exalted or debased by the new views of science. The nineteenth-century poet Arthur Hugh Clough (1819–1861) cries out in his poem “The New Sinai”:

> And as old from Sinai’s top God said that God is one,
> By science strict so speaks He now to tell us, there is None!
> Earth goes by chemic forces; Heaven’s a Mécanique Celeste!
> And heart and mind of humankind a watchwork as the rest!

(A “mécanique celeste” is a clockwork model to demonstrate celestial motions.)
The twentieth-century poet Robinson Jeffers (whose brother was an astronomer) saw it differently in a poem called “Star Swirls”:
"There is nothing like astronomy to pull the stuff out of man.
His stupid dreams and red-rooster importance:
Let him count the star-swirls."

Key Concepts and Summary

Calculating the gravitational interaction of more than two objects is complicated and requires large computers. If one object (like the Sun in our solar system) dominates gravitationally, it is possible to calculate the effects of a second object in terms of small perturbations. This approach was used by John Couch Adams and Urbain Le Verrier to predict the position of Neptune from its perturbations of the orbit of Uranus and thus discover a new planet mathematically.

For Further Exploration

Articles

Brahe and Kepler

Christianson, G. “The Celestial Palace of Tycho Brahe.” *Scientific American* (February 1961): 118.

Gingerich, O. “Johannes Kepler and the Rudolphine Tables.” *Sky & Telescope* (December 1971): 328. Brief article on Kepler’s work.

Wilson, C. “How Did Kepler Discover His First Two Laws?” *Scientific American* (March 1972): 92.

Newton

Christianson, G. “Newton’s *Principia*: A Retrospective.” *Sky & Telescope* (July 1987): 18.

Cohen, I. “Newton’s Discovery of Gravity.” *Scientific American* (March 1981): 166.

Gingerich, O. “Newton, Halley, and the Comet.” *Sky & Telescope* (March 1986): 230.

Sullivant, R. “When the Apple Falls.” *Astronomy* (April 1998): 55. Brief overview.

The Discovery of Neptune

Sheehan, W., et al. “The Case of the Pilfered Planet: Did the British Steal Neptune?” *Scientific American* (December 2004): 92.

Websites

Brahe and Kepler

Johannes Kepler: His Life, His Laws, and Time: http://kepler.nasa.gov/Mission/JohannesKepler/. From NASA’s Kepler mission.

Johannes Kepler: http://www.britannica.com/biography/Johannes-Kepler. Encyclopedia Britannica article.

Johannes Kepler: http://www-history.mcs.st-andrews.ac.uk/Biographies/Kepler.html. MacTutor article with additional links.

Noble Dane: Images of Tycho Brahe: http://www.mhs.ox.ac.uk/tycho/index.htm. A virtual museum exhibit from Oxford.

Newton

Sir Isaac Newton: http://www-groups.dcs.st-and.ac.uk/~history//Biographies/Newton.html. MacTutor article with additional links.

Sir Isaac Newton: http://www.luminarium.org/sevenlit/newton/newtonbio.htm. Newton Biography at the Luminarium.

The Discovery of Neptune

Adams, Airy, and the Discovery of Neptune: http://www.mikeoates.org/lassell/adams-airy.htm. A defense of Airy's role by historian Alan Chapman.

Mathematical Discovery of Planets: http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Neptune_and_Pluto.html. MacTutor article.

Videos

Brahe and Kepler

“Harmony of the Worlds.” This third episode of Carl Sagan’s TV series *Cosmos* focuses on Kepler and his life and work.

Tycho Brahe, Johannes Kepler, and Planetary Motion: https://www.youtube.com/watch?v=x3ALuycrCwI. German-produced video, in English (14:27).

Newton

Beyond the Big Bang: Sir Isaac Newton’s Law of Gravity: http://www.history.com/topics/enlightenment/videos/beyond-the-big-bang-sir-isaac-newtons-law-of-gravity. From the History Channel (4:35).

Sir Isaac Newton versus Bill Nye: Epic Rap Battles of History: https://www.youtube.com/watch?v=8yis7GzlXNM. (2:47).

The Discovery of Neptune

Richard Feynman: On the Discovery of Neptune: https://www.youtube.com/watch?v=FgXQffVgZRs. A brief black-and-white Caltech lecture (4:33).

Collaborative Group Activities

A. An eccentric, but very rich, alumnus of your college makes a bet with the dean that if you drop a baseball and a bowling ball from the tallest building on campus, the bowling ball would hit the ground first. Have your group discuss whether you would make a side bet that the alumnus is right. How would you decide who is right?
B. Suppose someone in your astronomy class was unhappy about his or her weight. Where could a person go to weigh one-fourth as much as he or she does now? Would changing the unhappy person’s weight have any effect on his or her mass?
C. When the Apollo astronauts landed on the Moon, some commentators commented that it ruined the mystery and “poetry” of the Moon

forever (and that lovers could never gaze at the full moon in the same way again). Others felt that knowing more about the Moon could only enhance its interest to us as we see it from Earth. How do the various members of your group feel? Why?

D. [link] shows a swarm of satellites in orbit around Earth. What do you think all these satellites do? How many categories of functions for Earth satellites can your group come up with?
E. The Making Connections feature box Astronomy and the Poets discusses how poets included the most recent astronomical knowledge in their poetry. Is this still happening today? Can your group members come up with any poems or songs that you know that deal with astronomy or outer space? If not, perhaps you could find some online, or by asking friends or roommates who are into poetry or music.

Review Questions

Exercise:

Problem: State Kepler's three laws in your own words.

Exercise:

Problem:

Why did Kepler need Tycho Brahe's data to formulate his laws?

Exercise:

Problem:

Which has more mass: an armful of feathers or an armful of lead? Which has more volume: a kilogram of feathers or a kilogram of lead? Which has higher density: a kilogram of feathers or a kilogram of lead?

Exercise:

Problem:

Explain how Kepler was able to find a relationship (his third law) between the orbital periods and distances of the planets that did not depend on the masses of the planets or the Sun.

Exercise:

Problem:

Write out Newton's three laws of motion in terms of what happens with the momentum of objects.

Exercise:

Problem: Which major planet has the largest . . .

A. semimajor axis?
B. average orbital speed around the Sun?
C. orbital period around the Sun?
D. eccentricity?

Exercise:

Problem:

Why do we say that Neptune was the first planet to be discovered through the use of mathematics?

Exercise:

Problem:

Why was Brahe reluctant to provide Kepler with all his data at one time?

Exercise:

Problem:

According to Kepler's second law, where in a planet's orbit would it be moving fastest? Where would it be moving slowest?

Exercise:

Problem:

The gas pedal, the brakes, and the steering wheel all have the ability to accelerate a car—how?

Exercise:

Problem:

Explain how a rocket can propel itself using Newton's third law.

Exercise:

Problem:

A certain material has a mass of 565 g while occupying 50 cm^3 of space. What is this material? (Hint: Use [link].)

Exercise:

Problem:

To calculate the momentum of an object, which properties of an object do you need to know?

Exercise:

Problem:

To calculate the angular momentum of an object, which properties of an object do you need to know?

Exercise:

Problem:

What was the great insight Newton had regarding Earth's gravity that allowed him to develop the universal law of gravitation?

Exercise:

Problem:

Which of these properties of an object best quantifies its inertia: velocity, acceleration, volume, mass, or temperature?

Exercise:

Problem:

Pluto's orbit is more eccentric than any of the major planets. What does that mean?

Exercise:

Problem:

Why is Tycho Brahe often called "the greatest naked-eye astronomer" of all time?

Thought Questions

Exercise:

Problem:

Is it possible to escape the force of gravity by going into orbit around Earth? How does the force of gravity in the International Space Station (orbiting an average of 400 km above Earth's surface) compare with that on the ground?

Exercise:

Problem:

What is the momentum of an object whose velocity is zero? How does Newton's first law of motion include the case of an object at rest?

Exercise:

Problem:

Evil space aliens drop you and your fellow astronomy student 1 km apart out in space, very far from any star or planet. Discuss the effects of gravity on each of you.

Exercise:

Problem:

A body moves in a perfectly circular path at constant speed. Are there forces acting in such a system? How do you know?

Exercise:

Problem:

As friction with our atmosphere causes a satellite to spiral inward, closer to Earth, its orbital speed increases. Why?

Exercise:

Problem:

Use a history book, an encyclopedia, or the internet to find out what else was happening in England during Newton's lifetime and discuss what trends of the time might have contributed to his accomplishments and the rapid acceptance of his work.

Exercise:

Problem:

Two asteroids begin to gravitationally attract one another. If one asteroid has twice the mass of the other, which one experiences the greater force? Which one experiences the greater acceleration?

Exercise:

Problem:

How does the mass of an astronaut change when she travels from Earth to the Moon? How does her weight change?

Exercise:

Problem:

If there is gravity where the International Space Station (ISS) is located above Earth, why doesn't the space station get pulled back down to Earth?

Exercise:

Problem:

Compare the density, weight, mass, and volume of a pound of gold to a pound of iron on the surface of Earth.

Exercise:

Problem:

If identical spacecraft were orbiting Mars and Earth at identical radii (distances), which spacecraft would be moving faster? Why?

Figuring for Yourself

Exercise:

Problem:

By what factor would a person's weight be increased if Earth had 10 times its present mass, but the same volume?

Exercise:

Problem:

Suppose astronomers find an earthlike planet that is twice the size of Earth (that is, its radius is twice that of Earth's). What must be the mass of this planet such that the gravitational force (F_{gravity}) at the surface would be identical to Earth's?

Exercise:

Problem:

What is the semimajor axis of a circle of diameter 24 cm? What is its eccentricity?

Exercise:

Problem:

If 24 g of material fills a cube 2 cm on a side, what is the density of the material?

Exercise:

Problem:

If 128 g of material is in the shape of a brick 2 cm wide, 4 cm high, and 8 cm long, what is the density of the material?

Exercise:

Problem:

If the major axis of an ellipse is 16 cm, what is the semimajor axis? If the eccentricity is 0.8, would this ellipse be best described as mostly circular or very elongated?

Exercise:

Problem:

What is the average distance from the Sun (in astronomical units) of an asteroid with an orbital period of 8 years?

Exercise:

Problem:

What is the average distance from the Sun (in astronomical units) of a planet with an orbital period of 45.66 years?

Exercise:

Problem:

In 1996, astronomers discovered an icy object beyond Pluto that was given the designation 1996 TL 66. It has a semimajor axis of 84 AU. What is its orbital period according to Kepler’s third law?

Glossary

perturbation
: a small disturbing effect on the motion or orbit of a body produced by a third body

Thinking Ahead
class="introduction"
Southern Summer.

As captured with a fish-eye lens aboard the Atlantis Space Shuttle on December 9, 1993, Earth hangs above the Hubble Space Telescope as it is repaired. The reddish continent is Australia, its size and shape distorted by the special lens. Because the seasons in the Southern Hemisphere are opposite those in the Northern Hemisphere, it is summer

in Australia on this December day. (credit: modification of work by NASA)

If Earth’s orbit is nearly a perfect circle (as we saw in earlier chapters), why is it hotter in summer and colder in winter in many places around the globe? And why are the seasons in Australia or Peru the opposite of those in the United States or Europe?

The story is told that Galileo, as he left the Hall of the Inquisition following his retraction of the doctrine that Earth rotates and revolves about the Sun, said under his breath, “But nevertheless it moves.” Historians are not sure whether the story is true, but certainly Galileo knew that Earth was in motion, whatever church authorities said.

It is the motions of Earth that produce the seasons and give us our measures of time and date. The Moon’s motions around us provide the concept of the month and the cycle of lunar phases. In this chapter we examine some of the basic phenomena of our everyday world in their astronomical context.

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe how latitude and longitude are used to map Earth
- Explain how right ascension and declination are used to map the sky

In order to create an accurate map, a mapmaker needs a way to uniquely and simply identify the location of all the major features on the map, such as cities or natural landmarks. Similarly, astronomical mapmakers need a way to uniquely and simply identify the location of stars, galaxies, and other celestial objects. On Earth maps, we divide the surface of Earth into a grid, and each location on that grid can easily be found using its *latitude* and *longitude* coordinate. Astronomers have a similar system for objects on the sky. Learning about these can help us understand the apparent motion of objects in the sky from various places on Earth.

Locating Places on Earth

Let's begin by fixing our position on the surface of planet Earth. As we discussed in Observing the Sky: The Birth of Astronomy, Earth's axis of rotation defines the locations of its North and South Poles and of its equator, halfway between. Two other directions are also defined by Earth's motions: east is the direction toward which Earth rotates, and west is its opposite. At almost any point on Earth, the four directions—north, south, east, and west—are well defined, despite the fact that our planet is round

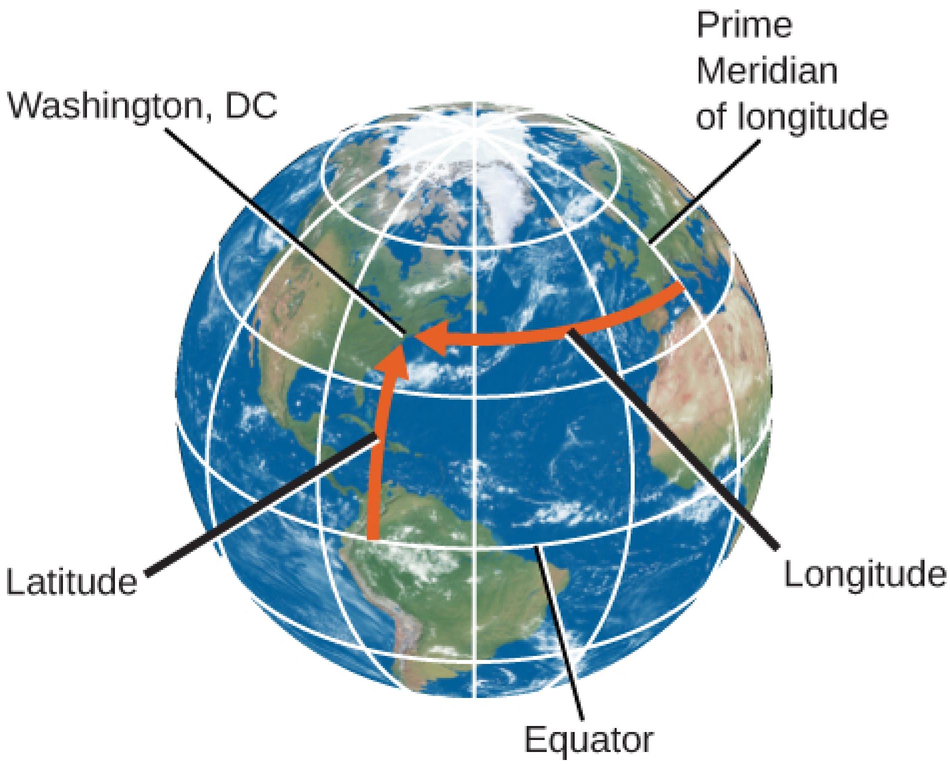
rather that flat. The only exceptions are exactly at the North and South Poles, where the directions east and west are ambiguous (because points exactly at the poles do not turn).

We can use these ideas to define a system of coordinates attached to our planet. Such a system, like the layout of streets and avenues in Manhattan or Salt Lake City, helps us find where we are or want to go. Coordinates on a sphere, however, are a little more complicated than those on a flat surface. We must define circles on the sphere that play the same role as the rectangular grid that you see on city maps.

A **great circle** is any circle on the surface of a sphere whose center is at the center of the sphere. For example, Earth's equator is a great circle on Earth's surface, halfway between the North and South Poles. We can also imagine a series of great circles that pass through both the North and South Poles. Each of the circles is called a **meridian**; they are each perpendicular to the equator, crossing it at right angles.

Any point on the surface of Earth will have a meridian passing through it ([link]). The meridian specifies the east-west location, or longitude, of the place. By international agreement (and it took many meetings for the world's countries to agree), longitude is defined as the number of degrees of arc along the equator between your meridian and the one passing through Greenwich, England, which has been designated as the Prime Meridian. The longitude of the Prime Meridian is defined as 0°.

Latitude and Longitude of Washington, DC.



We use latitude and longitude to find cities like Washington, DC, on a globe. Latitude is the number of degrees north or south of the equator, and longitude is the number of degrees east or west of the Prime Meridian. Washington, DC’s coordinates are 38° N and 77° W.

Why Greenwich, you might ask? Every country wanted 0° longitude to pass through its own capital. Greenwich, the site of the old Royal Observatory ([link]), was selected because it was between continental Europe and the United States, and because it was the site for much of the development of the method to measure longitude at sea. Longitudes are measured either to the east or to the west of the Greenwich meridian from 0° to 180°. As an example, the longitude of the clock-house benchmark of the U.S. Naval Observatory in Washington, DC, is 77.066° W.

Royal Observatory in Greenwich, England.

At the internationally agreed-upon zero point of longitude at the Royal Observatory Greenwich, tourists can stand and straddle the exact line where longitude “begins.”(credit left: modification of work by “pdbreen”/Flickr; credit right: modification of work by Ben Sutherland)

Your latitude (or north-south location) is the number of degrees of arc you are away from the equator along your meridian. Latitudes are measured either north or south of the equator from 0° to 90°. (The latitude of the equator is 0°.) As an example, the latitude of the previously mentioned Naval Observatory benchmark is 38.921° N. The latitude of the South Pole is 90° S, and the latitude of the North Pole is 90° N.

Locating Places in the Sky

Positions in the sky are measured in a way that is very similar to the way we measure positions on the surface of Earth. Instead of latitude and longitude, however, astronomers use coordinates called **declination** and **right ascension**. To denote positions of objects in the sky, it is often convenient to make use of the fictitious celestial sphere. We saw in Observing the Sky: The Birth of Astronomy that the sky appears to rotate about points above the North and South Poles of Earth—points in the sky called the north celestial pole and the south celestial pole. Halfway between

the celestial poles, and thus 90° from each pole, is the *celestial equator*, a great circle on the celestial sphere that is in the same plane as Earth’s equator. We can use these markers in the sky to set up a system of celestial coordinates.

Declination on the celestial sphere is measured the same way that latitude is measured on the sphere of Earth: from the celestial equator toward the north (positive) or south (negative). So Polaris, the star near the north celestial pole, has a declination of almost +90°.

Right ascension (RA) is like longitude, except that instead of Greenwich, the arbitrarily chosen point where we start counting is the *vernal equinox*, a point in the sky where the *ecliptic* (the Sun’s path) crosses the celestial equator. RA can be expressed either in units of angle (degrees) or in units of time. This is because the celestial sphere appears to turn around Earth once a day as our planet turns on its axis. Thus the 360° of RA that it takes to go once around the celestial sphere can just as well be set equal to 24 hours. Then each 15° of arc is equal to 1 hour of time. For example, the approximate celestial coordinates of the bright star Capella are RA 5h = 75° and declination +50°.

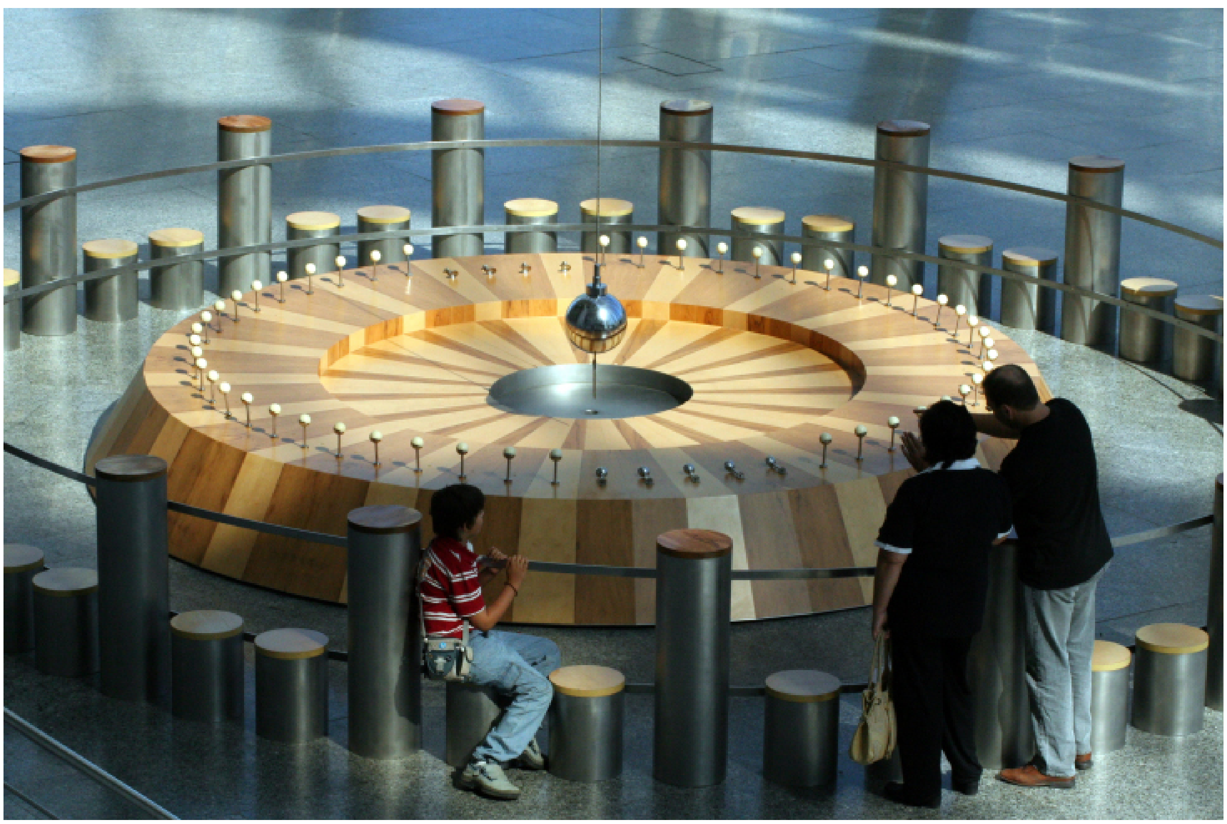
One way to visualize these circles in the sky is to imagine Earth as a transparent sphere with the terrestrial coordinates (latitude and longitude) painted on it with dark paint. Imagine the celestial sphere around us as a giant ball, painted white on the inside. Then imagine yourself at the center of Earth, with a bright light bulb in the middle, looking out through its transparent surface to the sky. The terrestrial poles, equator, and meridians will be projected as dark shadows on the celestial sphere, giving us the system of coordinates in the sky.

Note:
You can explore a variety of basic animations about coordinates and motions in the sky at this interactive site from ClassAction. Click on the “Animations” tab for a list of options. If you choose the second option in the menu, you can play with the celestial sphere and see RA and declination defined visually.

The Turning Earth

Why do many stars rise and set each night? Why, in other words, does the night sky seem to turn? We have seen that the apparent rotation of the celestial sphere could be accounted for either by a daily rotation of the sky around a stationary Earth or by the rotation of Earth itself. Since the seventeenth century, it has been generally accepted that it is Earth that turns, but not until the nineteenth century did the French physicist Jean Foucault provide an unambiguous demonstration of this rotation. In 1851, he suspended a 60-meter pendulum weighing about 25 kilograms from the dome of the Pantheon in Paris and started the pendulum swinging evenly. If Earth had not been turning, there would have been no alteration of the pendulum's plane of oscillation, and so it would have continued tracing the same path. Yet after a few minutes Foucault could see that the pendulum's plane of motion was turning. Foucault explained that it was not the pendulum that was shifting, but rather Earth that was turning beneath it ([link]). You can now find such pendulums in many science centers and planetariums around the world.

Foucault's Pendulum.



As Earth turns, the plane of oscillation of the Foucault pendulum shifts gradually so that over the course of 12 hours, all the targets in the circle at the edge of the wooden platform are knocked over in sequence. (credit: Manuel M. Vicente)

Can you think of other pieces of evidence that indicate that it is Earth and not the sky that is turning? (See Collaborative Group Activity A at the end of this chapter.)

Key Concepts and Summary

The terrestrial system of latitude and longitude makes use of the great circles called meridians. Longitude is arbitrarily set to 0° at the Royal Observatory at Greenwich, England. An analogous celestial coordinate system is called right ascension (RA) and declination, with 0° of declination starting at the vernal equinox. These coordinate systems help us locate any object on the celestial sphere. The Foucault pendulum is a way to demonstrate that Earth is turning.

Glossary

declination
: the angular distance north or south of the celestial equator

great circle
: a circle on the surface of a sphere that is the curve of intersection of the sphere with a plane passing through its center

meridian
: a great circle on the terrestrial or celestial sphere that passes through the poles

right ascension
: the coordinate for measuring the east-west positions of celestial bodies; the angle measured eastward along the celestial equator from

the vernal equinox to the hour circle passing through a body

Keeping Time

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the difference between the solar day and the sidereal day
- Explain mean solar time and the reason for time zones

The measurement of time is based on the rotation of Earth. Throughout most of human history, time has been reckoned by positions of the Sun and stars in the sky. Only recently have mechanical and electronic clocks taken over this function in regulating our lives.

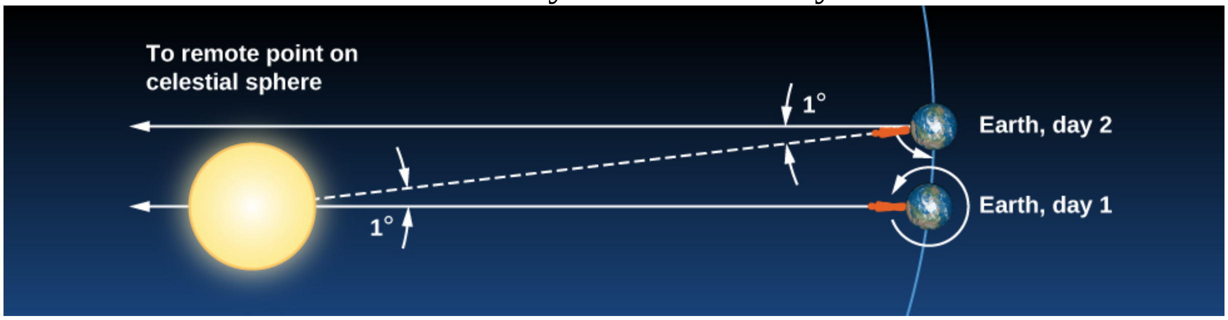
The Length of the Day

The most fundamental astronomical unit of time is the day, measured in terms of the rotation of Earth. There is, however, more than one way to define the day. Usually, we think of it as the rotation period of Earth with respect to the Sun, called the **solar day**. After all, for most people sunrise is more important than the rising time of Arcturus or some other star, so we set our clocks to some version of Sun-time. However, astronomers also use a **sidereal day**, which is defined in terms of the rotation period of Earth with respect to the stars.

A solar day is slightly longer than a sidereal day because (as you can see from [link]) Earth not only turns but also moves along its path around the Sun in a day. Suppose we start when Earth's orbital position is at day 1,

with both the Sun and some distant star (located in the direction indicated by the long white arrow pointing left), directly in line with the zenith for the observer on Earth. When Earth has completed one rotation with respect to the distant star and is at day 2, the long arrow again points to the same distant star. However, notice that because of the movement of Earth along its orbit from day 1 to 2, the Sun has not yet reached a position above the observer. To complete a solar day, Earth must rotate an additional amount, equal to 1/365 of a full turn. The time required for this extra rotation is 1/365 of a day, or about 4 minutes. So the solar day is about 4 minutes longer than the sidereal day.

Difference Between a Sidereal Day and a Solar Day.



This is a top view, looking down as Earth orbits the Sun. Because Earth moves around the Sun (roughly 1° per day), after one complete rotation of Earth relative to the stars, we do not see the Sun in the same position.

Because our ordinary clocks are set to solar time, stars rise 4 minutes earlier each day. Astronomers prefer sidereal time for planning their observations because in that system, a star rises at the same time every day.

Example:

Sidereal Time and Solar Time

The Sun makes a complete circle in the sky approximately every 24 hours, while the stars make a complete circle in the sky in 4 minutes less time, or 23 hours and 56 minutes. This causes the positions of the stars at a given

time of day or night to change slightly each day. Since stars rise 4 minutes earlier each day, that works out to about 2 hours per month (4 minutes × 30 = 120 minutes or 2 hours). So, if a particular constellation rises at sunset during the winter, you can be sure that by the summer, it will rise about 12 hours earlier, with the sunrise, and it will not be so easily visible in the night sky. Let’s say that tonight the bright star Sirius rises at 7:00 p.m. from a given location so that by midnight, it is very high in the sky. At what time will Sirius rise in three months?

Solution

In three months’ time, Sirius will be rising earlier by:

Equation:

$$90\text{ days } \times \frac{4\text{ minutes}}{\text{day}} = 360\text{ minutes or }6\text{ hours}$$

It will rise at about 1:00 p.m. and be high in the sky at around sunset instead of midnight. Sirius is the brightest star in the constellation of Canis Major (the big dog). So, some other constellation will be prominently visible high in the sky at this later date.

Check Your Learning

If a star rises at 8:30 p.m. tonight, approximately what time will it rise two months from now?

Note:

Answer:

In two months, the star will rise:

$60\text{ days } \times \frac{4\text{ minutes}}{\text{day}} = 240\text{ minutes or }4\text{ hours earlier.}$

This means it will rise at 4:30 p.m.

Apparent Solar Time

We can define **apparent solar time** as time reckoned by the actual position of the Sun in the sky (or, during the night, its position below the horizon). This is the kind of time indicated by sundials, and it probably represents the earliest measure of time used by ancient civilizations. Today, we adopt the middle of the night as the starting point of the day and measure time in hours elapsed since midnight.

During the first half of the day, the Sun has not yet reached the meridian (the great circle in the sky that passes through our zenith). We designate those hours as before midday (*ante meridiem*, or a.m.), before the Sun reaches the local meridian. We customarily start numbering the hours after noon over again and designate them by p.m. (*post meridiem*), after the Sun reaches the local meridian.

Although apparent solar time seems simple, it is not really very convenient to use. The exact length of an apparent solar day varies slightly during the year. The eastward progress of the Sun in its annual journey around the sky is not uniform because the speed of Earth varies slightly in its elliptical orbit. Another complication is that Earth's axis of rotation is not perpendicular to the plane of its revolution. Thus, apparent solar time does not advance at a uniform rate. After the invention of mechanical clocks that run at a uniform rate, it became necessary to abandon the apparent solar day as the fundamental unit of time.

Mean Solar Time and Standard Time

Instead, we can consider the **mean solar time**, which is based on the average value of the solar day over the course of the year. A mean solar day contains exactly 24 hours and is what we use in our everyday timekeeping. Although mean solar time has the advantage of progressing at a uniform rate, it is still inconvenient for practical use because it is determined by the position of the Sun. For example, noon occurs when the Sun is highest in the sky on the meridian (but not necessarily at the zenith). But because we live on a round Earth, the exact time of noon is different as you change your longitude by moving east or west.

If mean solar time were strictly observed, people traveling east or west would have to reset their watches continually as the longitude changed, just to read the local mean time correctly. For instance, a commuter traveling from Oyster Bay on Long Island to New York City would have to adjust the time on the trip through the East River tunnel because Oyster Bay time is actually about 1.6 minutes more advanced than that of Manhattan. (Imagine an airplane trip in which an obnoxious flight attendant gets on the intercom every minute, saying, “Please reset your watch for local mean time.”)

Until near the end of the nineteenth century, every city and town in the United States kept its own local mean time. With the development of railroads and the telegraph, however, the need for some kind of standardization became evident. In 1883, the United States was divided into four standard time zones (now six, including Hawaii and Alaska), each with one system of time within that zone.

By 1900, most of the world was using the system of 24 standardized global time zones. Within each zone, all places keep the same *standard time*, with the local mean solar time of a standard line of longitude running more or less through the middle of each zone. Now travelers reset their watches only when the time change has amounted to a full hour. Pacific standard time is 3 hours earlier than eastern standard time, a fact that becomes painfully obvious in California when someone on the East Coast forgets and calls you at 5:00 a.m.

Globally, almost all countries have adopted one or more standard time zones, although one of the largest nations, India, has settled on a half-zone, being 5.5 hours from Greenwich standard. Also, China officially uses only one time zone, so all the clocks in that country keep the same time. In Tibet, for example, the Sun rises while the clocks (which keep Beijing time) say it is midmorning already.

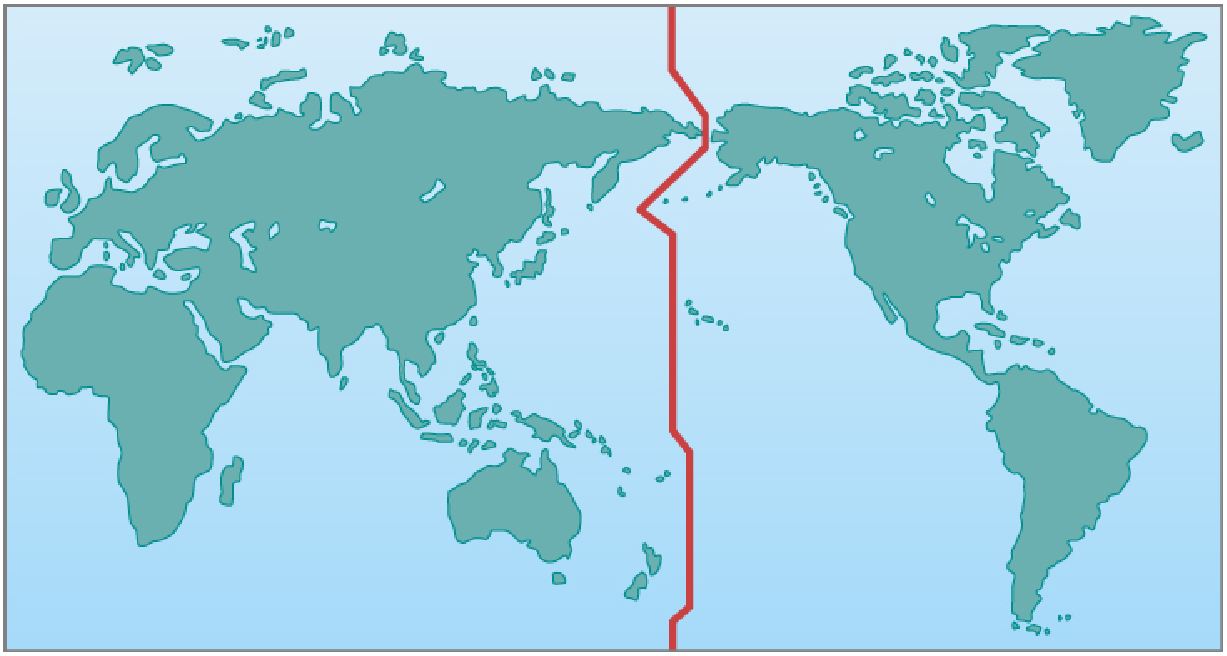
Daylight saving time is simply the local standard time of the place plus 1 hour. It has been adopted for spring and summer use in most states in the United States, as well as in many countries, to prolong the sunlight into evening hours, on the apparent theory that it is easier to change the time by government action than it would be for individuals or businesses to adjust their own schedules to produce the same effect. It does not, of course,

“save” any daylight at all—because the amount of sunlight is not determined by what we do with our clocks—and its observance is a point of legislative debate in some states.

The International Date Line

The fact that time is always advancing as you move toward the east presents a problem. Suppose you travel eastward around the world. You pass into a new time zone, on the average, about every 15° of longitude you travel, and each time you dutifully set your watch ahead an hour. By the time you have completed your trip, you have set your watch ahead a full 24 hours and thus gained a day over those who stayed at home.

The solution to this dilemma is the **International Date Line**, set by international agreement to run approximately along the 180° meridian of longitude. The date line runs down the middle of the Pacific Ocean, although it jogs a bit in a few places to avoid cutting through groups of islands and through Alaska ([link]). By convention, at the date line, the date of the calendar is changed by one day. Crossing the date line from west to east, thus advancing your time, you compensate by decreasing the date; crossing from east to west, you increase the date by one day. To maintain our planet on a rational system of timekeeping, we simply must accept that the date will differ in different cities at the same time. A good example is the date when the Imperial Japanese Navy bombed Pearl Harbor in Hawaii, known in the United States as Sunday, December 7, 1941, but taught to Japanese students as Monday, December 8.
Where the Date Changes.



The International Date Line is an arbitrarily drawn line on Earth where the date changes. So that neighbors do not have different days, the line is located where Earth's surface is mostly water.

Key Concepts and Summary

The basic unit of astronomical time is the day—either the solar day (reckoned by the Sun) or the sidereal day (reckoned by the stars). Apparent solar time is based on the position of the Sun in the sky, and mean solar time is based on the average value of a solar day during the year. By international agreement, we define 24 time zones around the world, each with its own standard time. The convention of the International Date Line is necessary to reconcile times on different parts of Earth.

Glossary

apparent solar time
: time as measured by the position of the Sun in the sky (the time that would be indicated by a sundial)

International Date Line

an arbitrary line on the surface of Earth near longitude 180° across which the date changes by one day

mean solar time

time based on the rotation of Earth; mean solar time passes at a constant rate, unlike apparent solar time

sidereal day

Earth’s rotation period as defined by the positions of the stars in the sky; the time between successive passages of the same star through the meridian

solar day

Earth’s rotation period as defined by the position of the Sun in the sky; the time between successive passages of the Sun through the meridian

The Calendar

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Understand how calendars varied among different cultures
- Explain the origins of our modern calendar

“What’s today’s date?” is one of the most common questions you can ask (usually when signing a document or worrying about whether you should have started studying for your next astronomy exam). Long before the era of digital watches, smartphones, and fitness bands that tell the date, people used calendars to help measure the passage of time.

The Challenge of the Calendar

There are two traditional functions of any calendar. First, it must keep track of time over the course of long spans, allowing people to anticipate the cycle of the seasons and to honor special religious or personal anniversaries. Second, to be useful to a large number of people, a calendar must use natural time intervals that everyone can agree on—those defined by the motions of Earth, the Moon, and sometimes even the planets. The natural units of our calendar are the *day*, based on the period of rotation of Earth; the *month*, based on the cycle of the Moon’s phases (see later in this chapter) about Earth; and the year, based on the period of revolution of Earth about the Sun. Difficulties have resulted from the fact that these three periods are not commensurable; that’s a fancy way of saying that one does not divide evenly into any of the others.

The rotation period of Earth is, by definition, 1.0000 day (and here the solar day is used, since that is the basis of human experience). The period required by the Moon to complete its cycle of phases, called the *month*, is 29.5306 days. The basic period of revolution of Earth, called the *tropical year*, is 365.2422 days. The ratios of these numbers are not convenient for calculations. This is the historic challenge of the calendar, dealt with in various ways by different cultures.

Early Calendars

Even the earliest cultures were concerned with the keeping of time and the calendar. Some interesting examples include monuments left by Bronze Age people in northwestern Europe, especially the British Isles. The best preserved of the monuments is Stonehenge, about 13 kilometers from Salisbury in southwest England ([link]). It is a complex array of stones, ditches, and holes arranged in concentric circles. Carbon dating and other studies show that Stonehenge was built during three periods ranging from about 2800 to 1500 BCE. Some of the stones are aligned with the directions of the Sun and Moon during their risings and settings at critical times of the year (such as the summer and winter solstices), and it is generally believed that at least one function of the monument was connected with the keeping of a calendar.

Stonehenge.

The ancient monument known as Stonehenge was used to keep track of the motions of the Sun and Moon. (credit: modification of work by Adriano Aurelio Araujo)

The Maya in Central America, who thrived more than a thousand years ago, were also concerned with the keeping of time. Their calendar was as sophisticated as, and perhaps more complex than, contemporary calendars in Europe. The Maya did not attempt to correlate their calendar accurately with the length of the year or lunar month. Rather, their calendar was a system for keeping track of the passage of days and for counting time far into the past or future. Among other purposes, it was useful for predicting astronomical events, such as the position of Venus in the sky ([link]).
El Caracol.

This Mayan observatory at Chichen Itza in the Yucatan, Mexico, dates from around the year 1000. (credit: “wiredtourist.com”/Flickr)

The ancient Chinese developed an especially complex calendar, largely limited to a few privileged hereditary court astronomer-astrologers. In addition to the motions of Earth and the Moon, they were able to fit in the approximately 12-year cycle of Jupiter, which was central to their system of astrology. The Chinese still preserve some aspects of this system in their cycle of 12 “years”—the Year of the Dragon, the Year of the Pig, and so on—that are defined by the position of Jupiter in the zodiac.

Our Western calendar derives from a long history of timekeeping beginning with the Sumerians, dating back to at least the second millennium BCE, and continuing with the Egyptians and the Greeks around the eighth century BCE. These calendars led, eventually, to the *Julian calendar*, introduced by Julius Caesar, which approximated the year at 365.25 days, fairly close to the actual value of 365.2422. The Romans achieved this approximation by declaring years to have 365 days each, with the exception of every fourth year. The *leap year* was to have one extra day, bringing its length to 366

days, and thus making the average length of the year in the Julian calendar 365.25 days.

In this calendar, the Romans had dropped the almost impossible task of trying to base their calendar on the Moon as well as the Sun, although a vestige of older lunar systems can be seen in the fact that our months have an average length of about 30 days. However, lunar calendars remained in use in other cultures, and Islamic calendars, for example, are still primarily lunar rather than solar.

The Gregorian Calendar

Although the Julian calendar (which was adopted by the early Christian Church) represented a great advance, its average year still differed from the true year by about 11 minutes, an amount that accumulates over the centuries to an appreciable error. By 1582, that 11 minutes per year had added up to the point where the first day of spring was occurring on March 11, instead of March 21. If the trend were allowed to continue, eventually the Christian celebration of Easter would be occurring in early winter. Pope Gregory XIII, a contemporary of Galileo, felt it necessary to institute further calendar reform.

The Gregorian calendar reform consisted of two steps. First, 10 days had to be dropped out of the calendar to bring the vernal equinox back to March 21; by proclamation, the day following October 4, 1582, became October 15. The second feature of the new Gregorian calendar was a change in the rule for leap year, making the average length of the year more closely approximate the tropical year. Gregory decreed that three of every four century years—all leap years under the Julian calendar—would be common years henceforth. The rule was that only century years divisible by 400 would be leap years. Thus, 1700, 1800, and 1900—all divisible by 4 but not by 400—were not leap years in the Gregorian calendar. On the other hand, the years 1600 and 2000, both divisible by 400, were leap years. The average length of this Gregorian year, 365.2425 mean solar days, is correct to about 1 day in 3300 years.

The Catholic countries immediately put the Gregorian reform into effect, but countries of the Eastern Church and most Protestant countries did not adopt it until much later. It was 1752 when England and the American colonies finally made the change. By parliamentary decree, September 2, 1752, was followed by September 14. Although special laws were passed to prevent such abuses as landlords collecting a full month's rent for September, there were still riots, and people demanded their 12 days back. Russia did not abandon the Julian calendar until the time of the Bolshevik revolution. The Russians then had to omit 13 days to come into step with the rest of the world. The anniversary of the October Revolution (old calendar) of 1917, bringing the communists to power, thus ended up being celebrated in November (new calendar), a difference that is perhaps not so important since the fall of communism.

Key Concepts and Summary

The fundamental problem of the calendar is to reconcile the incommensurable lengths of the day, month, and year. Most modern calendars, beginning with the Roman (Julian) calendar of the first century BCE, neglect the problem of the month and concentrate on achieving the correct number of days in a year by using such conventions as the leap year. Today, most of the world has adopted the Gregorian calendar established in 1582 while finding ways to coexist with the older lunar calendars' system of months.

Ocean Tides and the Moon

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe what causes tides on Earth
- Explain why the amplitude of tides changes during the course of a month

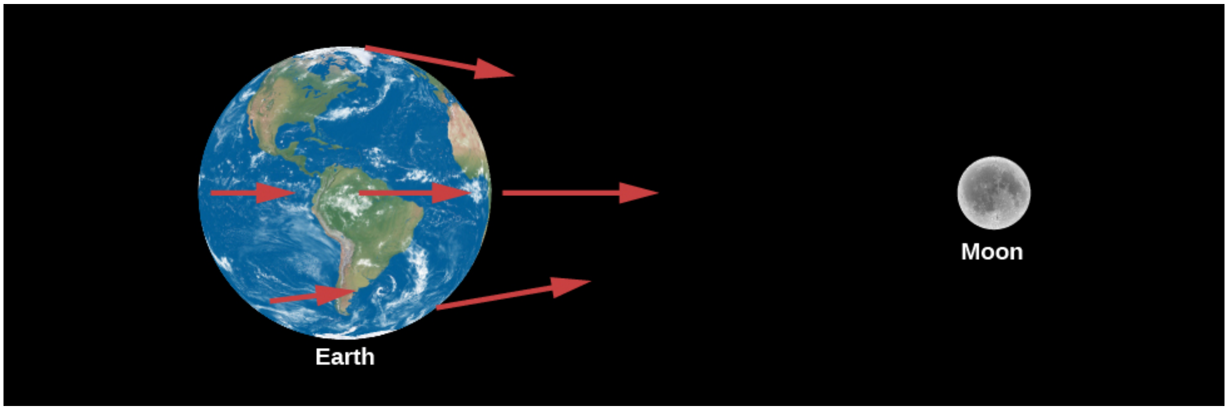
Anyone living near the sea is familiar with the twice-daily rising and falling of the **tides**. Early in history, it was clear that tides must be related to the Moon because the daily delay in high tide is the same as the daily delay in the Moon’s rising. A satisfactory explanation of the tides, however, awaited the theory of gravity, supplied by Newton.

The Pull of the Moon on Earth

The gravitational forces exerted by the Moon at several points on Earth are illustrated in [link]. These forces differ slightly from one another because Earth is not a point, but has a certain size: all parts are not equally distant from the Moon, nor are they all in exactly the same direction from the Moon. Moreover, Earth is not perfectly rigid. As a result, the differences among the forces of the Moon’s attraction on different parts of Earth (called *differential forces*) cause Earth to distort slightly. The side of Earth nearest the Moon is attracted toward the Moon more strongly than is the center of Earth, which in turn is attracted more strongly than is the side opposite the Moon. Thus, the differential forces tend to stretch Earth slightly into a

prolate spheroid (a football shape), with its long diameter pointed toward the Moon.

Pull of the Moon.



The Moon’s differential attraction is shown on different parts of Earth. (Note that the differences have been exaggerated for educational purposes.)

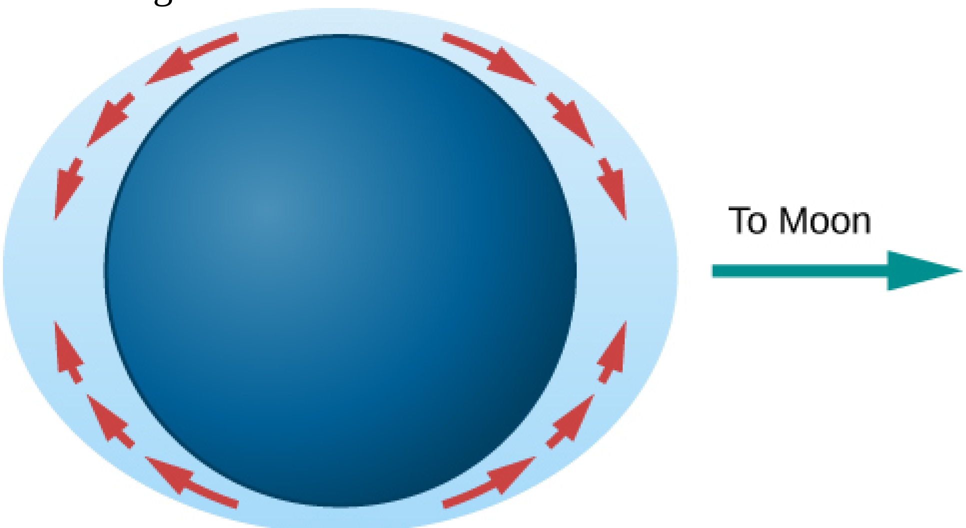
If Earth were made of water, it would distort until the Moon’s differential forces over different parts of its surface came into balance with Earth’s own gravitational forces pulling it together. Calculations show that in this case, Earth would distort from a sphere by amounts ranging up to nearly 1 meter. Measurements of the actual deformation of Earth show that the solid Earth does distort, but only about one-third as much as water would, because of the greater rigidity of Earth’s interior.

Because the tidal distortion of the solid Earth amounts—at its greatest—to only about 20 centimeters, Earth does not distort enough to balance the Moon’s differential forces with its own gravity. Hence, objects at Earth’s surface experience tiny horizontal tugs, tending to make them slide about. These *tide-raising forces* are too insignificant to affect solid objects like astronomy students or rocks in Earth’s crust, but they do affect the waters in the oceans.

The Formation of Tides

The tide-raising forces, acting over a number of hours, produce motions of the water that result in measurable tidal bulges in the oceans. Water on the side of Earth facing the Moon flows toward it, with the greatest depths roughly at the point below the Moon. On the side of Earth opposite the Moon, water also flows to produce a tidal bulge ([link]).

Tidal Bulges in an “Ideal” Ocean.



Differences in gravity cause tidal forces that push water in the direction of tidal bulges on Earth.

Note:
You can run this animation for a visual demonstration of the tidal bulge.

Note that the tidal bulges in the oceans do not result from the Moon’s compressing or expanding the water, nor from the Moon’s lifting the water “away from Earth.” Rather, they result from an actual flow of water over Earth’s surface toward the two regions below and opposite the Moon, causing the water to pile up to greater depths at those places ([link]).

High and Low Tides.

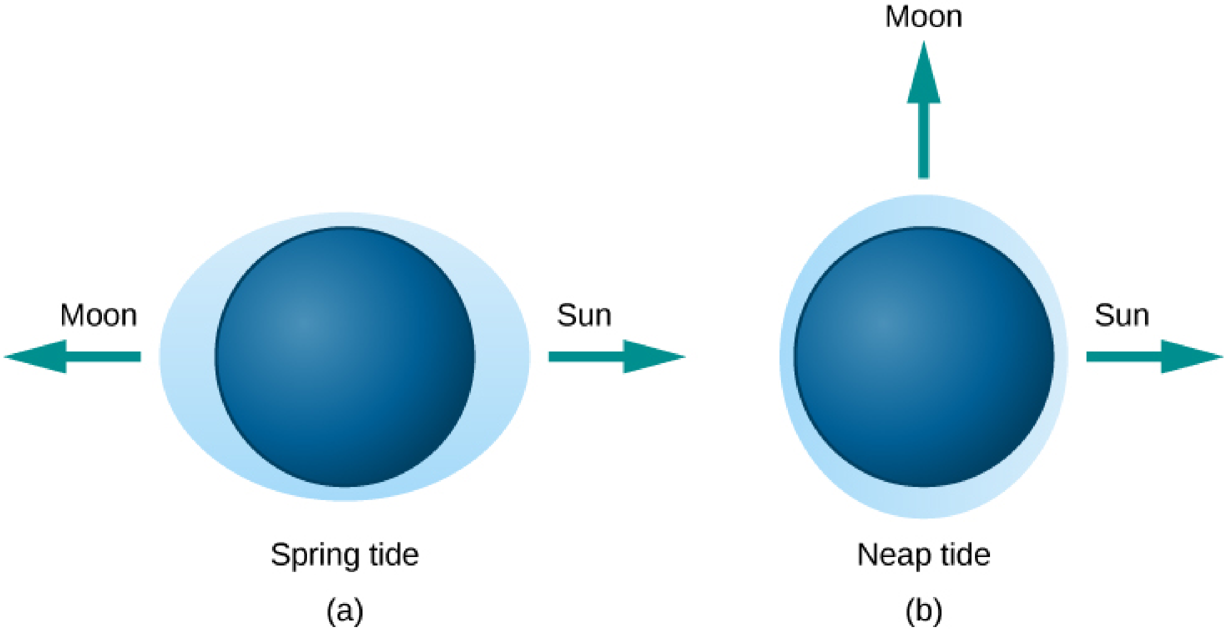
This is a side-by-side comparison of the Bay of Fundy in Canada at high and low tides. (credit a, b: modification of work by Dylan Kereluk)

In the idealized (and, as we shall see, oversimplified) model just described, the height of the tides would be only a few feet. The rotation of Earth would carry an observer at any given place alternately into regions of deeper and shallower water. An observer being carried toward the regions under or opposite the Moon, where the water was deepest, would say, “The tide is coming in”; when carried away from those regions, the observer would say, “The tide is going out.” During a day, the observer would be carried through two tidal bulges (one on each side of Earth) and so would experience two high tides and two low tides.

The Sun also produces tides on Earth, although it is less than half as effective as the Moon at tide raising. The actual tides we experience are a combination of the larger effect of the Moon and the smaller effect of the Sun. When the Sun and Moon are lined up (at new moon or full moon), the tides produced reinforce each other and so are greater than normal ([link]). These are called spring tides (the name is connected not to the season but to the idea that higher tides “spring up”). Spring tides are approximately the same, whether the Sun and Moon are on the same or opposite sides of Earth, because tidal bulges occur on both sides. When the Moon is at first quarter or last quarter (at right angles to the Sun’s direction), the tides

produced by the Sun partially cancel the tides of the Moon, making them lower than usual. These are called neap tides.

Tides Caused by Different Alignments of the Sun and Moon.



(a) In spring tides, the Sun's and Moon's pulls reinforce each other. (b) In neap tides, the Sun and the Moon pull at right angles to each other and the resulting tides are lower than usual.

The "simple" theory of tides, described in the preceding paragraphs, would be sufficient if Earth rotated very slowly and were completely surrounded by very deep oceans. However, the presence of land masses stopping the flow of water, the friction in the oceans and between oceans and the ocean floors, the rotation of Earth, the wind, the variable depth of the ocean, and other factors all complicate the picture. This is why, in the real world, some places have very small tides while in other places huge tides become tourist attractions. If you have been in such places, you may know that "tide tables" need to be computed and published for each location; one set of tide predictions doesn't work for the whole planet. In this introductory chapter, we won't delve further into these complexities.

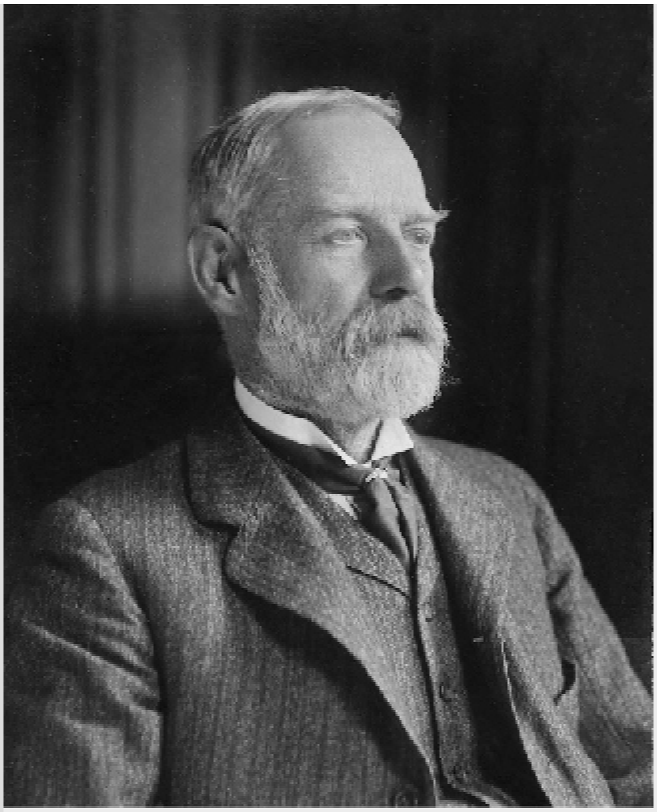
Note:

George Darwin and the Slowing of Earth

The rubbing of water over the face of Earth involves an enormous amount of energy. Over long periods of time, the friction of the tides is slowing down the rotation of Earth. Our day gets longer by about 0.002 second each century. That seems very small, but such tiny changes can add up over millions and billions of years.

Although Earth's spin is slowing down, the angular momentum (see Orbits and Gravity) in a system such as the Earth-Moon system cannot change. Thus, some other spin motion must speed up to take the extra angular momentum. The details of what happens were worked out over a century ago by George Darwin, the son of naturalist Charles Darwin. George Darwin (see [link]) had a strong interest in science but studied law for six years and was admitted to the bar. However, he never practiced law, returning to science instead and eventually becoming a professor at Cambridge University. He was a protégé of Lord Kelvin, one of the great physicists of the nineteenth century, and he became interested in the long-term evolution of the solar system. He specialized in making detailed (and difficult) mathematical calculations of how orbits and motions change over geologic time.

George Darwin (1845–1912).



George Darwin is best known for studying Earth’s spin in relation to angular momentum.

What Darwin calculated for the Earth-Moon system was that the Moon will slowly spiral outward, away from Earth. As it moves farther away, it will orbit less quickly (just as planets farther from the Sun move more slowly in their orbits). Thus, the month will get longer. Also, because the Moon will be more distant, total eclipses of the Sun will no longer be visible from Earth.

Both the day and the month will continue to get longer, although bear in mind that the effects are very gradual. Darwin’s calculations were confirmed by mirrors placed on the Moon by Apollo 11 astronauts. These show that the Moon is moving away by 3.8 centimeters per year, and that ultimately—billions of years in the future—the day and the month will be the same length (about 47 of our present days). At this point the Moon will be stationary in the sky over the same spot on Earth, meaning some parts

of Earth will see the Moon and its phases and other parts will never see them. This kind of alignment is already true for Pluto’s moon Charon (among others). Its rotation and orbital period are the same length as a day on Pluto.

Key Concepts and Summary

The twice-daily ocean tides are primarily the result of the Moon’s differential force on the material of Earth’s crust and ocean. These tidal forces cause ocean water to flow into two tidal bulges on opposite sides of Earth; each day, Earth rotates through these bulges. Actual ocean tides are complicated by the additional effects of the Sun and by the shape of the coasts and ocean basins.

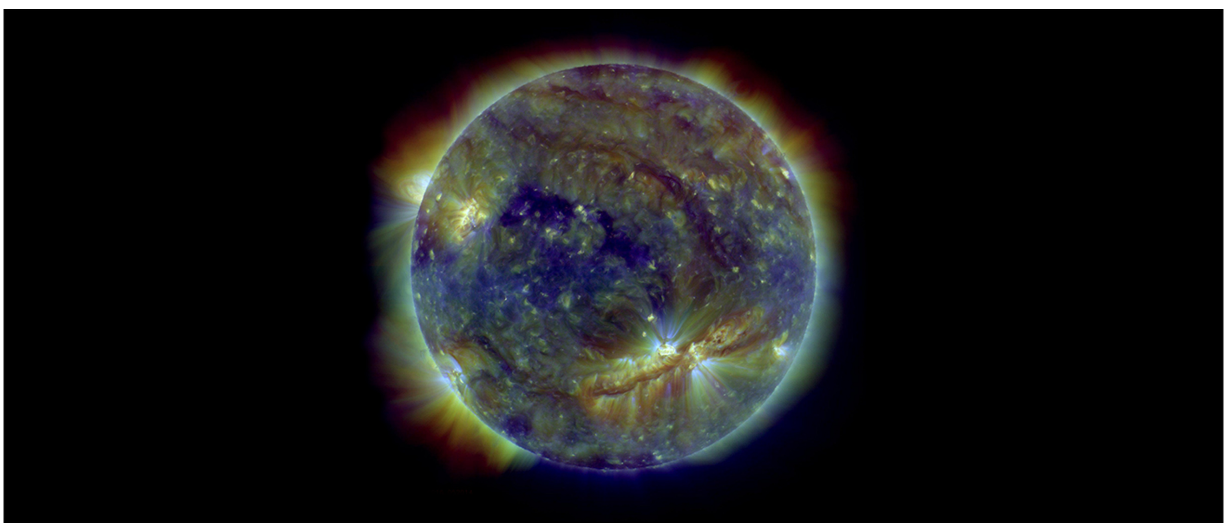
Glossary

tides
: alternate rising and falling of sea level caused by the difference in the strength of the Moon’s gravitational pull on different parts of Earth

Thinking Ahead
class="introduction"
Our Sun in Ultraviolet Light.

This photograph of the Sun was taken at several different wavelengths of ultraviolet, which our eyes cannot see, and then color coded so it reveals activity in our Sun’s atmosphere that cannot be observed in visible light. This is why it is important to observe the Sun and other astronomica l objects in wavelengths other than the visible band of the

spectrum. This image was taken by a satellite from above Earth’s atmosphere, which is necessary since Earth’s atmosphere absorbs much of the ultraviolet light coming from space. (credit: modification of work by NASA)



The nearest star is so far away that the fastest spacecraft humans have built would take almost 100,000 years to get there. Yet we very much want to know what material this neighbor star is composed of and how it differs

from our own Sun. How can we learn about the chemical makeup of stars that we cannot hope to visit or sample?

In astronomy, most of the objects that we study are completely beyond our reach. The temperature of the Sun is so high that a spacecraft would be fried long before it reached it, and the stars are much too far away to visit in our lifetimes with the technology now available. Even light, which travels at a speed of 300,000 kilometers per second (km/s), takes more than 4 years to reach us from the nearest star. If we want to learn about the Sun and stars, we must rely on techniques that allow us to analyze them from a distance.

The Behavior of Light

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the evidence for Maxwell’s electromagnetic model of light
- Describe the relationship between wavelength, frequency, and speed of light
- Discuss the particle model of light and the definition of photon
- Explain how and why the amount of light we see from an object depends upon its distance

Coded into the light and other kinds of radiation that reach us from objects in the universe is a wide range of information about what those objects are like and how they work. If we can decipher this code and read the messages it contains, we can learn an enormous amount about the cosmos without ever having to leave Earth or its immediate environment.

The visible light and other radiation we receive from the stars and planets is generated by processes at the atomic level—by changes in the way the parts of an atom interact and move. Thus, to appreciate how light is generated, we must explore how atoms work. There is a bit of irony in the fact that in order to understand some of the largest structures in the universe, we must become acquainted with some of the smallest.

Notice that we have twice used the phrase “light and other radiation.” One of the key ideas explored in this chapter is that visible light is not unique; it

is merely the most familiar example of a much larger family of radiation that can carry information to us.

The word “radiation” will be used frequently in this book, so it is important to understand what it means. In everyday language, “radiation” is often used to describe certain kinds of energetic subatomic particles released by radioactive materials in our environment. (An example is the kind of radiation used to treat some cancers.) But this is not what we mean when we use the word “radiation” in an astronomy text. *Radiation*, as used in this book, is a general term for waves (including light waves) that *radiate* outward from a source.

As we saw in Orbits and Gravity, Newton’s theory of gravity accounts for the motions of planets as well as objects on Earth. Application of this theory to a variety of problems dominated the work of scientists for nearly two centuries. In the nineteenth century, many physicists turned to the study of electricity and magnetism, which are intimately connected with the production of light.

The scientist who played a role in this field comparable to Newton’s role in the study of gravity was physicist James Clerk Maxwell, born and educated in Scotland ([link]). Inspired by a number of ingenious experiments that showed an intimate relationship between electricity and magnetism, Maxwell developed a theory that describes both electricity and magnetism with only a small number of elegant equations. It is this theory that gives us important insights into the nature and behavior of light.

James Clerk Maxwell (1831–1879).

Maxwell unified the rules governing electricity and magnetism into a coherent theory.

Maxwell’s Theory of Electromagnetism

We will look at the structure of the atom in more detail later, but we begin by noting that the typical atom consists of several types of particles, a number of which have not only mass but an additional property called electric charge. In the nucleus (central part) of every atom are *protons*, which are positively charged; outside the nucleus are electrons, which have a negative charge.

Maxwell’s theory deals with these electric charges and their effects, especially when they are moving. In the vicinity of an electron charge, another charge feels a force of attraction or repulsion: opposite charges attract; like charges repel. When charges are not in motion, we observe only this electric attraction or repulsion. If charges are in motion, however (as

they are inside every atom and in a wire carrying a current), then we measure another force called *magnetism*.

Magnetism was well known for much of recorded human history, but its cause was not understood until the nineteenth century. Experiments with electric charges demonstrated that magnetism was the result of moving charged particles. Sometimes, the motion is clear, as in the coils of heavy wire that make an industrial electromagnet. Other times, it is more subtle, as in the kind of magnet you buy in a hardware store, in which many of the electrons inside the atoms are spinning in roughly the same direction; it is the alignment of their motion that causes the material to become magnetic.

Physicists use the word *field* to describe the action of forces that one object exerts on other distant objects. For example, we say the Sun produces a *gravitational field* that controls Earth's orbit, even though the Sun and Earth do not come directly into contact. Using this terminology, we can say that stationary electric charges produce *electric fields*, and moving electric charges also produce *magnetic fields*.

Actually, the relationship between electric and magnetic phenomena is even more profound. Experiments showed that changing magnetic fields could produce electric currents (and thus changing electric fields), and changing electric currents could in turn produce changing magnetic fields. So once begun, electric and magnetic field changes could continue to trigger each other.

Maxwell analyzed what would happen if electric charges were oscillating (moving constantly back and forth) and found that the resulting pattern of electric and magnetic fields would spread out and travel rapidly through space. Something similar happens when a raindrop strikes the surface of water or a frog jumps into a pond. The disturbance moves outward and creates a pattern we call a *wave* in the water ([link]). You might, at first, think that there must be very few situations in nature where electric charges oscillate, but this is not at all the case. As we shall see, atoms and molecules (which consist of charged particles) oscillate back and forth all the time. The resulting electromagnetic disturbances are among the most common phenomena in the universe.

Making Waves.

An oscillation in a pool of water creates an expanding disturbance called a wave. (credit: modification of work by "vastateparksstaff"/Flickr)

Maxwell was able to calculate the speed at which an electromagnetic disturbance moves through space; he found that it is equal to the speed of light, which had been measured experimentally. On that basis, he speculated that light was one form of a family of possible electromagnetic disturbances called **electromagnetic radiation**, a conclusion that was again confirmed in laboratory experiments. When light (reflected from the pages of an astronomy textbook, for example) enters a human eye, its changing electric and magnetic fields stimulate nerve endings, which then transmit the information contained in these changing fields to the brain. The science of astronomy is primarily about analyzing radiation from distant objects to understand what they are and how they work.

The Wave-Like Characteristics of Light

The changing electric and magnetic fields in light are similar to the waves that can be set up in a quiet pool of water. In both cases, the disturbance

travels rapidly outward from the point of origin and can use its energy to disturb other things farther away. (For example, in water, the expanding ripples moving away from our frog could disturb the peace of a dragonfly resting on a leaf in the same pool.) In the case of electromagnetic waves, the radiation generated by a transmitting antenna full of charged particles and moving electrons at your local radio station can, sometime later, disturb a group of electrons in your car radio antenna and bring you the news and weather while you are driving to class or work in the morning.

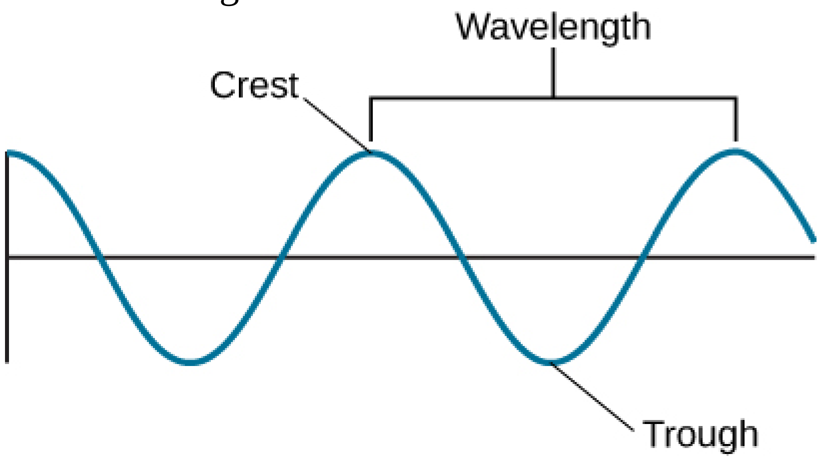
The waves generated by charged particles differ from water waves in some profound ways, however. Water waves require water to travel in. The sound waves we hear, to give another example, are pressure disturbances that require air to travel though. But electromagnetic waves do not require water or air: the fields generate each other and so can move through a vacuum (such as outer space). This was such a disturbing idea to nineteenth-century scientists that they actually made up a substance to fill all of space—one for which there was not a single shred of evidence—just so light waves could have something to travel through: they called it the *aether*. Today, we know that there is no aether and that electromagnetic waves have no trouble at all moving through empty space (as all the starlight visible on a clear night must surely be doing).

The other difference is that *all* electromagnetic waves move at the same speed in empty space (the speed of light—approximately 300,000 kilometers per second, or 300,000,000 meters per second, which can also be written as 3×10^8 m/s), which turns out to be the fastest possible speed in the universe. No matter where electromagnetic waves are generated from and no matter what other properties they have, when they are moving (and not interacting with matter), they move at the speed of light. Yet you know from everyday experience that there are different kinds of light. For example, we perceive that light waves differ from one another in a property we call color. Let’s see how we can denote the differences among the whole broad family of electromagnetic waves.

The nice thing about a wave is that it is a repeating phenomenon. Whether it is the up-and-down motion of a water wave or the changing electric and magnetic fields in a wave of light, the pattern of disturbance repeats in a

cyclical way. Thus, any wave motion can be characterized by a series of crests and troughs ([link]). Moving from one crest through a trough to the next crest completes one cycle. The horizontal length covered by one cycle is called the **wavelength**. Ocean waves provide an analogy: the wavelength is the distance that separates successive wave crests.

Characterizing Waves.



Electromagnetic radiation has wave-like characteristics. The wavelength (λ) is the distance between crests, the frequency (*f*) is the number of cycles per second, and the speed (*c*) is the distance the wave covers during a specified period of time (e.g., kilometers per second).

For visible light, our eyes perceive different wavelengths as different colors: red, for example, is the longest visible wavelength, and violet is the shortest. The main colors of visible light from longest to shortest wavelength can be remembered using the mnemonic ROY G BIV—for Red, Orange, Yellow, Green, Blue, Indigo, and Violet. Other invisible forms of electromagnetic radiation have different wavelengths, as we will see in the next section.

We can also characterize different waves by their **frequency**, the number of wave cycles that pass by per second. If you count 10 crests moving by each

second, for example, then the frequency is 10 cycles per second (cps). In honor of Heinrich Hertz, the physicist who—inspired by Maxwell's work—discovered radio waves, a cps is also called a *hertz* (Hz). Take a look at your radio, for example, and you will see the channel assigned to each radio station is characterized by its frequency, usually in units of KHz (kilohertz, or thousands of hertz) or MHz (megahertz, or millions of hertz).

Wavelength (λ) and frequency (*f*) are related because all electromagnetic waves travel at the same speed. To see how this works, imagine a parade in which everyone is forced by prevailing traffic conditions to move at exactly the same speed. You stand on a corner and watch the waves of marchers come by. First you see row after row of miniature ponies. Because they are not very large and, therefore, have a shorter wavelength, a good number of the ponies can move past you each minute; we can say they have a high frequency. Next, however, come several rows of circus elephants. The elephants are large and marching at the same speed as the ponies, so far fewer of them can march past you per minute: Because they have a wider spacing (longer wavelength), they represent a lower frequency.

The formula for this relationship can be expressed as follows: for any wave motion, the speed at which a wave moves equals the frequency times the wavelength. Waves with longer wavelengths have lower frequencies. Mathematically, we can express this as
Equation:

$$c = \lambda f$$

where the Greek letter for "l"—lambda, λ—is used to denote wavelength and *c* is the scientific symbol for the speed of light. Solving for the wavelength, this is expressed as:
Equation:

$$\lambda = \frac{c}{f}.$$

Example:

Deriving and Using the Wave Equation

The equation for the relationship between the speed and other characteristics of a wave can be derived from our basic understanding of motion. The average speed of anything that is moving is:

Equation:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}$$

(So, for example, a car on the highway traveling at a speed of 100 km/h covers 100 km during the time of 1 h.) For an electromagnetic wave to travel the distance of one of its wavelengths, λ, at the speed of light, *c*, we have $c = \lambda/t$. The frequency of a wave is the number of cycles per second. If a wave has a frequency of a million cycles per second, then the time for each cycle to go by is a millionth of a second. So, in general, $t = 1/f$. Substituting into our wave equation, we get $c = \lambda \times f$. Now let's use this to calculate an example. What is the wavelength of visible light that has a frequency of 5.66×10^{14} Hz?

Solution

Solving the wave equation for wavelength, we find:

Equation:

$$\lambda = \frac{c}{f}$$

Substituting our values gives:

Equation:

$$\lambda = \frac{3.00 \times 10^{8}\,\text{m/s}}{5.66 \times 10^{14}\,\text{Hz}} = 5.30 \times 10^{-7}\,\text{m}$$

This answer can also be written as 530 nm, which is in the yellow-green part of the visible spectrum (nm stands for nanometers, where the term "nano" means "billionths").

Check Your Learning

"Tidal waves," or tsunamis, are waves caused by earthquakes that travel rapidly through the ocean. If a tsunami travels at the speed of 600 km/h

and approaches a shore at a rate of one wave crest every 15 min (4 waves/h), what would be the distance between those wave crests at sea?

Note:
Answer:
$$\lambda = \frac{600\ \text{km/h}}{4\ \text{waves/h}} = 150\ \text{km}$$

Light as a Photon

The electromagnetic wave model of light (as formulated by Maxwell) was one of the great triumphs of nineteenth-century science. In 1887, when Heinrich Hertz actually made invisible electromagnetic waves (what today are called radio waves) on one side of a room and detected them on the other side, it ushered in a new era that led to the modern age of telecommunications. His experiment ultimately led to the technologies of television, cell phones, and today's wireless networks around the globe.

However, by the beginning of the twentieth century, more sophisticated experiments had revealed that light behaves in certain ways that cannot be explained by the wave model. Reluctantly, physicists had to accept that sometimes light behaves more like a "particle"—or at least a self-contained packet of energy—than a wave. We call such a packet of electromagnetic energy a **photon**.

The fact that light behaves like a wave in certain experiments and like a particle in others was a very surprising and unlikely idea. After all, our common sense says that waves and particles are opposite concepts. On one hand, a wave is a repeating disturbance that, by its very nature, is not in only one place, but spreads out. A particle, on the other hand, is something that can be in only one place at any given time. Strange as it sounds,

though, countless experiments now confirm that electromagnetic radiation can sometimes behave like a wave and at other times like a particle.

Then, again, perhaps we shouldn't be surprised that something that always travels at the "speed limit" of the universe and doesn't need a medium to travel through might not obey our everyday common sense ideas. The confusion that this wave-particle duality of light caused in physics was eventually resolved by the introduction of a more complicated theory of waves and particles, now called quantum mechanics. (This is one of the most interesting fields of modern science, but it is mostly beyond the scope of our book. If you are interested in it, see some of the suggested resources at the end of this chapter.)

In any case, you should now be prepared when scientists (or the authors of this book) sometimes discuss electromagnetic radiation as if it consisted of waves and at other times refer to it as a stream of photons. A photon (being a packet of energy) carries a specific amount of energy. We can use the idea of energy to connect the photon and wave models. How much energy a photon has depends on its frequency when you think about it as a wave. A low-energy radio wave has a low frequency as a wave, while a high-energy X-ray at your dentist's office is a high-frequency wave. Among the colors of visible light, violet-light photons have the highest energy and red-light photons have the lowest.

Test whether the connection between photons and waves is clear to you. In the above example, which photon would have the longer wavelength as a wave: the radio wave or the X-ray? If you answered the radio wave, you are correct. Radio waves have a lower frequency, so the wave cycles are longer (they are elephants, not miniature ponies).

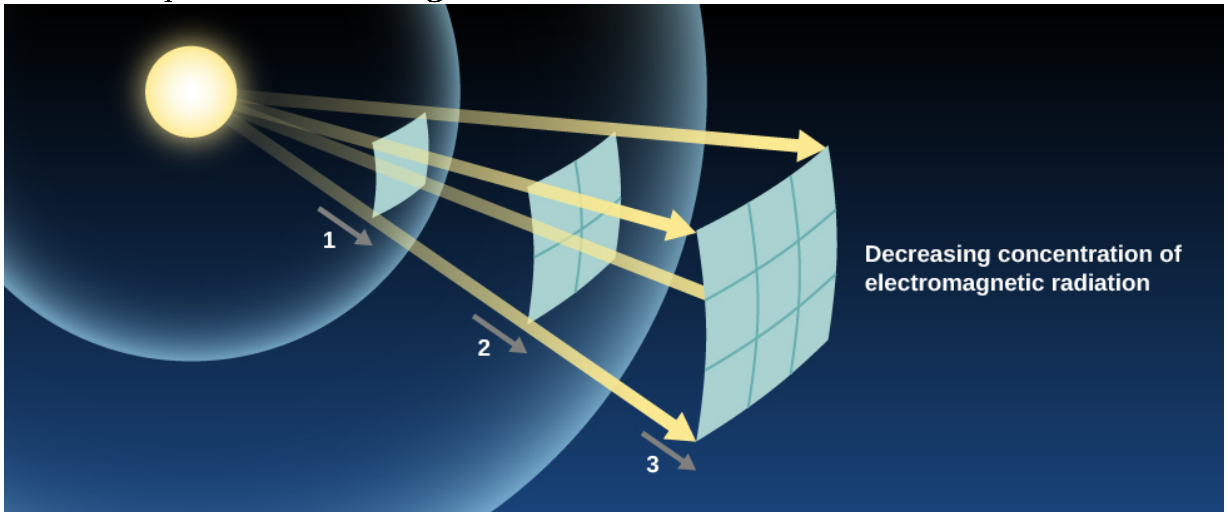
Propagation of Light

Let's think for a moment about how light from a lightbulb moves through space. As waves expand, they travel away from the bulb, not just toward your eyes but in all directions. They must therefore cover an ever-widening space. Yet the total amount of light available can't change once the light has left the bulb. This means that, as the same expanding shell of light covers a

larger and larger area, there must be less and less of it in any given place. Light (and all other electromagnetic radiation) gets weaker and weaker as it gets farther from its source.

The increase in the area that the light must cover is proportional to the square of the distance that the light has traveled ([link]). If we stand twice as far from the source, our eyes will intercept two-squared (2 × 2), or four times less light. If we stand 10 times farther from the source, we get 10-squared, or 100 times less light. You can see how this weakening means trouble for sources of light at astronomical distances. One of the nearest stars, Alpha Centauri A, emits about the same total energy as the Sun. But it is about 270,000 times farther away, and so it appears about 73 billion times fainter. No wonder the stars, which close-up would look more or less like the Sun, look like faint pinpoints of light from far away.

Inverse Square Law for Light.



As light radiates away from its source, it spreads out in such a way that the energy per unit area (the amount of energy passing through one of the small squares) decreases as the square of the distance from its source.

This idea—that the apparent brightness of a source (how bright it looks to us) gets weaker with distance in the way we have described—is known as the **inverse square law** for light propagation. In this respect, the

propagation of light is similar to the effects of gravity. Remember that the force of gravity between two attracting masses is also inversely proportional to the square of their separation.

Example:
The Inverse Square Law for Light
The intensity of a 120-W lightbulb observed from a distance 2 m away is 2.4 W/m^2. What would be the intensity if this distance was doubled?
Solution
If we move twice as far away, then the answer will change according to the inverse square of the distance, so the new intensity will be $(1/2)^2 = 1/4$ of the original intensity, or 0.6 W/m^2.
Check Your Learning
How many times brighter or fainter would a star appear if it were moved to:

a. twice its present distance?
b. ten times its present distance?
c. half its present distance?

Note:
Answer:
a. $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$; b. $\left(\frac{1}{10}\right)^2 = \frac{1}{100}$; c. $\left(\frac{1}{1/2}\right)^2 = 4$

Key Concepts and Summary

James Clerk Maxwell showed that whenever charged particles change their motion, as they do in every atom and molecule, they give off waves of

energy. Light is one form of this electromagnetic radiation. The wavelength of light determines the color of visible radiation. Wavelength (λ) is related to frequency (f) and the speed of light (c) by the equation $c = \lambda f$. Electromagnetic radiation sometimes behaves like waves, but at other times, it behaves as if it were a particle—a little packet of energy, called a photon. The apparent brightness of a source of electromagnetic energy decreases with increasing distance from that source in proportion to the square of the distance—a relationship known as the inverse square law.

Glossary

electromagnetic radiation
: radiation consisting of waves propagated through regularly varying electric and magnetic fields and traveling at the speed of light

frequency
: the number of waves that cross a given point per unit time (in radiation)

inverse square law
: (for light) the amount of energy (light) flowing through a given area in a given time decreases in proportion to the square of the distance from the source of energy or light

photon
: a discrete unit (or “packet”) of electromagnetic energy

wavelength
: the distance from crest to crest or trough to trough in a wave

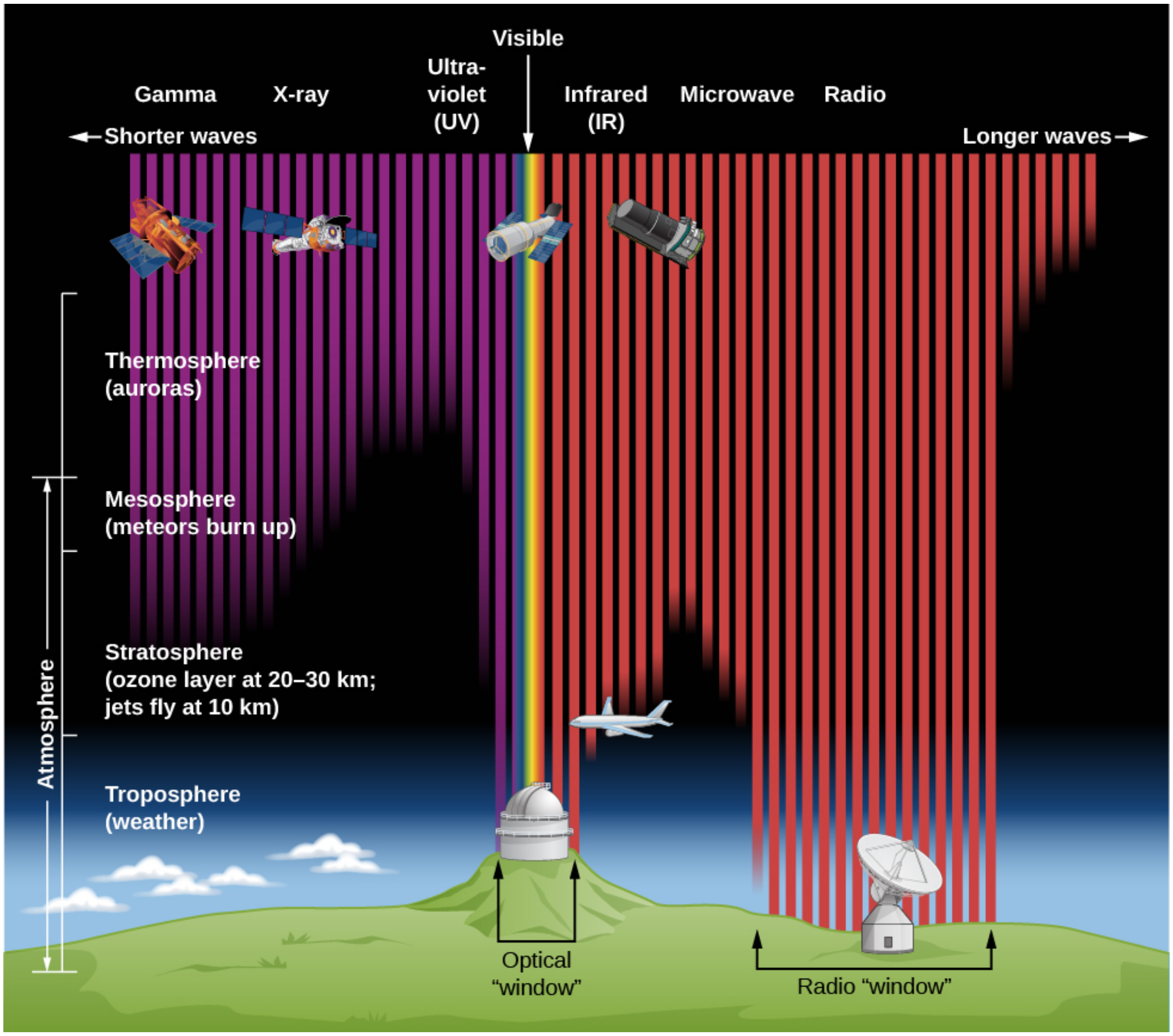
The Electromagnetic Spectrum

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Understand the bands of the electromagnetic spectrum and how they differ from one another
- Understand how each part of the spectrum interacts with Earth's atmosphere
- Explain how and why the light emitted by an object depends on its temperature

Objects in the universe send out an enormous range of electromagnetic radiation. Scientists call this range the **electromagnetic spectrum**, which they have divided into a number of categories. The spectrum is shown in [link], with some information about the waves in each part or band.
Radiation and Earth's Atmosphere.



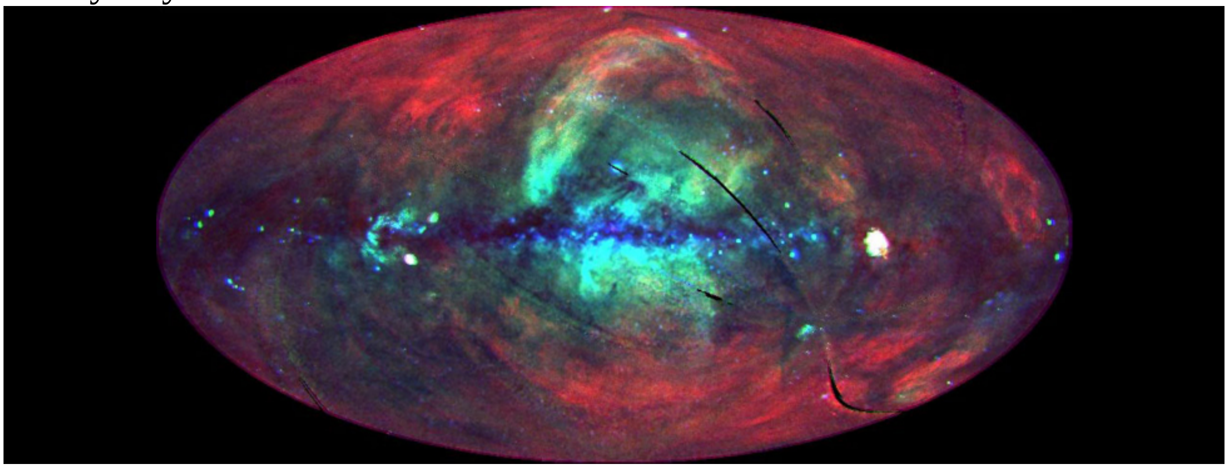
This figure shows the bands of the electromagnetic spectrum and how well Earth's atmosphere transmits them. Note that high-frequency waves from space do not make it to the surface and must therefore be observed from space. Some infrared and microwaves are absorbed by water and thus are best observed from high altitudes. Low-frequency radio waves are blocked by Earth's ionosphere. (credit: modification of work by STScI/JHU/NASA)

Types of Electromagnetic Radiation

Electromagnetic radiation with the shortest wavelengths, no longer than 0.01 nanometer, is categorized as **gamma rays** (1 nanometer = 10^{-9} meters; see Appendix D). The name *gamma* comes from the third letter of the Greek alphabet: gamma rays were the third kind of radiation discovered coming from radioactive atoms when physicists first investigated their behavior. Because gamma rays carry a lot of energy, they can be dangerous for living tissues. Gamma radiation is generated deep in the interior of stars, as well as by some of the most violent phenomena in the universe, such as the deaths of stars and the merging of stellar corpses. Gamma rays coming to Earth are absorbed by our atmosphere before they reach the ground (which is a good thing for our health); thus, they can only be studied using instruments in space.

Electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers is referred to as **X-rays**. Being more energetic than visible light, X-rays are able to penetrate soft tissues but not bones, and so allow us to make images of the shadows of the bones inside us. While X-rays can penetrate a short length of human flesh, they are stopped by the large numbers of atoms in Earth's atmosphere with which they interact. Thus, X-ray astronomy (like gamma-ray astronomy) could not develop until we invented ways of sending instruments above our atmosphere ([link]).

X-Ray Sky.



This is a map of the sky tuned to certain types of X-rays (seen from above Earth's atmosphere). The map tilts the sky so that the disk of our Milky Way Galaxy runs across its center. It was constructed and artificially colored from data gathered by the European ROSAT

satellite. Each color (red, yellow, and blue) shows X-rays of different frequencies or energies. For example, red outlines the glow from a hot local bubble of gas all around us, blown by one or more exploding stars in our cosmic vicinity. Yellow and blue show more distant sources of X-rays, such as remnants of other exploded stars or the active center of our Galaxy (in the middle of the picture). (credit: modification of work by NASA)

Radiation intermediate between X-rays and visible light is **ultraviolet** (meaning higher energy than violet). Outside the world of science, ultraviolet light is sometimes called “black light” because our eyes cannot see it. Ultraviolet radiation is mostly blocked by the ozone layer of Earth’s atmosphere, but a small fraction of ultraviolet rays from our Sun do penetrate to cause sunburn or, in extreme cases of overexposure, skin cancer in human beings. Ultraviolet astronomy is also best done from space.

Electromagnetic radiation with wavelengths between roughly 400 and 700 nm is called **visible light** because these are the waves that human vision can perceive. This is also the band of the electromagnetic spectrum that most readily reaches Earth’s surface. These two observations are not coincidental: human eyes evolved to see the kinds of waves that arrive from the Sun most effectively. Visible light penetrates Earth’s atmosphere effectively, except when it is temporarily blocked by clouds.

Between visible light and radio waves are the wavelengths of **infrared** or heat radiation. Astronomer William Herschel first discovered infrared in 1800 while trying to measure the temperatures of different colors of sunlight spread out into a spectrum. He noticed that when he accidently positioned his thermometer beyond the reddest color, it still registered heating due to some invisible energy coming from the Sun. This was the first hint about the existence of the other (invisible) bands of the electromagnetic spectrum, although it would take many decades for our full understanding to develop.

A heat lamp radiates mostly infrared radiation, and the nerve endings in our skin are sensitive to this band of the electromagnetic spectrum. Infrared waves are absorbed by water and carbon dioxide molecules, which are more concentrated low in Earth’s atmosphere. For this reason, infrared astronomy is best done from high mountaintops, high-flying airplanes, and spacecraft.

After infrared comes the familiar **microwave**, used in short-wave communication and microwave ovens. (Wavelengths vary from 1 millimeter to 1 meter and are absorbed by water vapor, which makes them effective in heating foods.) The “micro-” prefix refers to the fact that microwaves are small in comparison to radio waves, the next on the spectrum. You may remember that tea—which is full of water—heats up quickly in your microwave oven, while a ceramic cup—from which water has been removed by baking—stays cool in comparison.

All electromagnetic waves longer than microwaves are called **radio waves**, but this is so broad a category that we generally divide it into several subsections. Among the most familiar of these are radar waves, which are used in radar guns by traffic officers to determine vehicle speeds, and AM radio waves, which were the first to be developed for broadcasting. The wavelengths of these different categories range from over a meter to hundreds of meters, and other radio radiation can have wavelengths as long as several kilometers.

With such a wide range of wavelengths, not all radio waves interact with Earth’s atmosphere in the same way. FM and TV waves are not absorbed and can travel easily through our atmosphere. AM radio waves are absorbed or reflected by a layer in Earth’s atmosphere called the ionosphere (the ionosphere is a layer of charged particles at the top of our atmosphere, produced by interactions with sunlight and charged particles that are ejected from the Sun).

We hope this brief survey has left you with one strong impression: although visible light is what most people associate with astronomy, the light that our eyes can see is only a tiny fraction of the broad range of waves generated in the universe. Today, we understand that judging some astronomical phenomenon by using only the light we can see is like hiding under the table at a big dinner party and judging all the guests by nothing but their

shoes. There’s a lot more to each person than meets our eye under the table. It is very important for those who study astronomy today to avoid being “visible light chauvinists”—to respect only the information seen by their eyes while ignoring the information gathered by instruments sensitive to other bands of the electromagnetic spectrum.

[link] summarizes the bands of the electromagnetic spectrum and indicates the temperatures and typical astronomical objects that emit each kind of electromagnetic radiation. While at first, some of the types of radiation listed in the table may seem unfamiliar, you will get to know them better as your astronomy course continues. You can return to this table as you learn more about the types of objects astronomers study.

Types of Electromagnetic Radiation

Type of Radiation	Wavelength Range (nm)	Radiated by Objects at This Temperature	Typical Sources
Gamma rays	Less than 0.01	More than 10^8 K	Produced in nuclear reactions; require very high-energy processes
X-rays	0.01–20	10^6–10^8 K	Gas in clusters of galaxies, supernova remnants, solar corona

Types of Electromagnetic Radiation			
Type of Radiation	**Wavelength Range (nm)**	**Radiated by Objects at This Temperature**	**Typical Sources**
Ultraviolet	20–400	10^4–10^6 K	Supernova remnants, very hot stars
Visible	400–700	10^3–10^4 K	Stars
Infrared	10^3–10^6	10–10^3 K	Cool clouds of dust and gas, planets, moons
Microwave	10^6–10^9	Less than 10 K	Active galaxies, pulsars, cosmic background radiation
Radio	More than 10^9	Less than 10 K	Supernova remnants, pulsars, cold gas

Radiation and Temperature

Some astronomical objects emit mostly infrared radiation, others mostly visible light, and still others mostly ultraviolet radiation. What determines the type of electromagnetic radiation emitted by the Sun, stars, and other dense astronomical objects? The answer often turns out to be their *temperature*.

At the microscopic level, everything in nature is in motion. A solid is composed of molecules and atoms in continuous vibration: they move back and forth in place, but their motion is much too small for our eyes to make out. A gas consists of atoms and/or molecules that are flying about freely at high speed, continually bumping into one another and bombarding the surrounding matter. The hotter the solid or gas, the more rapid the motion of its molecules or atoms. The temperature of something is thus a measure of the average motion energy of the particles that make it up.

This motion at the microscopic level is responsible for much of the electromagnetic radiation on Earth and in the universe. As atoms and molecules move about and collide, or vibrate in place, their electrons give off electromagnetic radiation. The characteristics of this radiation are determined by the temperature of those atoms and molecules. In a hot material, for example, the individual particles vibrate in place or move rapidly from collisions, so the emitted waves are, on average, more energetic. And recall that higher energy waves have a higher frequency. In very cool material, the particles have low-energy atomic and molecular motions and thus generate lower-energy waves.

Note:
Check out the NASA briefing or NASA's 5-minute introductory video to learn more about the electromagnetic spectrum.

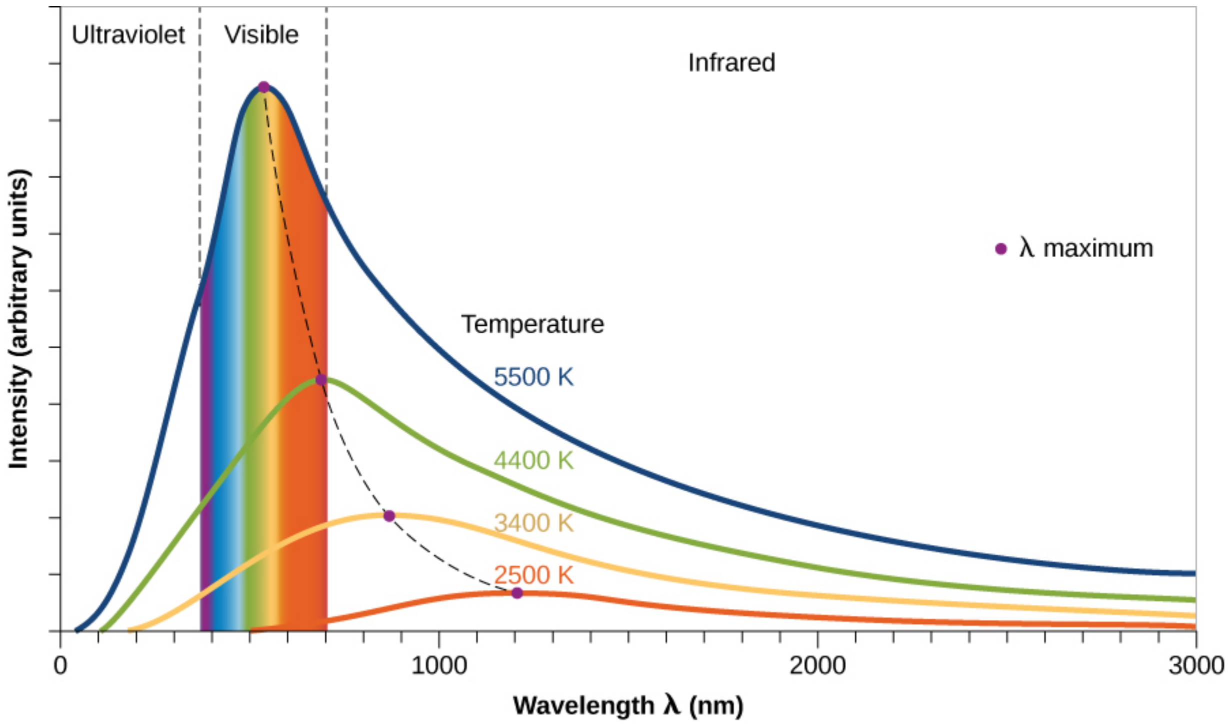
Radiation Laws

To understand, in more quantitative detail, the relationship between temperature and electromagnetic radiation, we imagine an idealized object called a **blackbody**. Such an object (unlike your sweater or your astronomy instructor's head) does not reflect or scatter any radiation, but absorbs all the electromagnetic energy that falls onto it. The energy that is absorbed causes the atoms and molecules in it to vibrate or move around at increasing speeds. As it gets hotter, this object will radiate electromagnetic waves until absorption and radiation are in balance. We want to discuss such an

idealized object because, as you will see, stars behave in very nearly the same way.

The radiation from a blackbody has several characteristics, as illustrated in [link]. The graph shows the power emitted at each wavelength by objects of different temperatures. In science, the word *power* means the energy coming off per second (and it is typically measured in *watts*, which you are probably familiar with from buying lightbulbs).

Radiation Laws Illustrated.



This graph shows in arbitrary units how many photons are given off at each wavelength for objects at four different temperatures. The wavelengths corresponding to visible light are shown by the colored bands. Note that at hotter temperatures, more energy (in the form of photons) is emitted at all wavelengths. The higher the temperature, the shorter the wavelength at which the peak amount of energy is radiated (this is known as Wien's law).

First of all, notice that the curves show that, at each temperature, our blackbody object emits radiation (photons) at all wavelengths (all colors).

This is because in any solid or denser gas, some molecules or atoms vibrate or move between collisions slower than average and some move faster than average. So when we look at the electromagnetic waves emitted, we find a broad range, or spectrum, of energies and wavelengths. More energy is emitted at the average vibration or motion rate (the highest part of each curve), but if we have a large number of atoms or molecules, some energy will be detected at each wavelength.

Second, note that an object at a higher temperature emits more power at all wavelengths than does a cooler one. In a hot gas (the taller curves in [link]), for example, the atoms have more collisions and give off more energy. In the real world of stars, this means that hotter stars give off more energy at every wavelength than do cooler stars.

Third, the graph shows us that the higher the temperature, the shorter the wavelength at which the maximum power is emitted. Remember that a shorter wavelength means a higher frequency and energy. It makes sense, then, that hot objects give off a larger fraction of their energy at shorter wavelengths (higher energies) than do cool objects. You may have observed examples of this rule in everyday life. When a burner on an electric stove is turned on low, it emits only heat, which is infrared radiation, but does not glow with visible light. If the burner is set to a higher temperature, it starts to glow a dull red. At a still-higher setting, it glows a brighter orange-red (shorter wavelength). At even higher temperatures, which cannot be reached with ordinary stoves, metal can appear brilliant yellow or even blue-white.

We can use these ideas to come up with a rough sort of “thermometer” for measuring the temperatures of stars. Because many stars give off most of their energy in visible light, the color of light that dominates a star’s appearance is a rough indicator of its temperature. If one star looks red and another looks blue, which one has the higher temperature? Because blue is the shorter-wavelength color, it is the sign of a hotter star. (Note that the temperatures we associate with different colors in science are not the same as the ones artists use. In art, red is often called a “hot” color and blue a “cool” color. Likewise, we commonly see red on faucet or air conditioning controls to indicate hot temperatures and blue to indicate cold temperatures.

Although these are common uses to us in daily life, in nature, it’s the other way around.)

We can develop a more precise star thermometer by measuring how much energy a star gives off at each wavelength and by constructing diagrams like [link]. The location of the peak (or maximum) in the power curve of each star can tell us its temperature. The average temperature at the surface of the Sun, which is where the radiation that we see is emitted, turns out to be 5800 K. (Throughout this text, we use the kelvin or absolute temperature scale. On this scale, water freezes at 273 K and boils at 373 K. All molecular motion ceases at 0 K. The various temperature scales are described in Appendix D.) There are stars cooler than the Sun and stars hotter than the Sun.

The wavelength at which maximum power is emitted can be calculated according to the equation
Equation:

$$\lambda_{\text{max}} = \frac{3 \times 10^6}{T}$$

where the wavelength is in nanometers (one billionth of a meter) and the temperature is in K (the constant 3 x 10^6 has units of nm × K). This relationship is called **Wien’s law**. For the Sun, the wavelength at which the maximum energy is emitted is 520 nanometers, which is near the middle of that portion of the electromagnetic spectrum called visible light. Characteristic temperatures of other astronomical objects, and the wavelengths at which they emit most of their power, are listed in [link].

Example:
Calculating the Temperature of a Blackbody
We can use Wien’s law to calculate the temperature of a star provided we know the wavelength of peak intensity for its spectrum. If the emitted radiation from a red dwarf star has a wavelength of maximum power at 1200 nm, what is the temperature of this star, assuming it is a blackbody?

Solution

Solving Wien's law for temperature gives:

Equation:

$$T=\frac{3\ \times\ 10^{6}\,\text{nm K}}{\lambda_{\text{max}}}=\frac{3\ \times\ 10^{6}\,\text{nm K}}{1200\ \text{nm}}=2500\,\text{K}$$

Check Your Learning

What is the temperature of a star whose maximum light is emitted at a much shorter wavelength of 290 nm?

Note:

Answer:

$T=\frac{3\times10^{6}\,\text{nm K}}{\lambda_{\text{max}}}=\frac{3\times10^{6}\,\text{nm K}}{290\ \text{nm}}=10{,}300\,\text{K}$

Since this star has a peak wavelength that is at a shorter wavelength (in the ultraviolet part of the spectrum) than that of our Sun (in the visible part of the spectrum), it should come as no surprise that its surface temperature is much hotter than our Sun's.

We can also describe our observation that hotter objects radiate more power at all wavelengths in a mathematical form. If we sum up the contributions from all parts of the electromagnetic spectrum, we obtain the total energy emitted by a blackbody. What we usually measure from a large object like a star is the **energy flux**, the power emitted per square meter. The word *flux* means "flow" here: we are interested in the flow of power into an area (like the area of a telescope mirror). It turns out that the energy flux from a blackbody at temperature *T* is proportional to the fourth power of its absolute temperature. This relationship is known as the **Stefan-Boltzmann law** and can be written in the form of an equation as

Equation:

$$F = \sigma T^4$$

where F stands for the energy flux (in units of watts per square meter), T is given in Kelvins, and σ (Greek letter sigma) is a constant number (5.67×10^{-8}).

Notice how impressive this result is. Increasing the temperature of a star would have a tremendous effect on the power it radiates. If the Sun, for example, were twice as hot—that is, if it had a temperature of 11,600 K—it would radiate 2^4, or 16 times more power than it does now. Tripling the temperature would raise the power output 81 times. Hot stars really shine away a tremendous amount of energy.

Example:

Calculating the Power of a Star

While energy flux tells us how much power a star emits per square meter, we would often like to know how much total power is emitted by the star. We can determine that by multiplying the energy flux by the number of square meters on the surface of the star. Stars are mostly spherical, so we can use the formula $4\pi R^2$ for the surface area, where R is the radius of the star. The total power emitted by the star (which we call the star's "absolute luminosity") can be found by multiplying the formula for energy flux and the formula for the surface area:

Equation:

$$L = 4\pi R^2 \sigma T^4$$

Two stars have the same size and are the same distance from us. Star A has a surface temperature of 6000 K, and star B has a surface temperature twice as high, 12,000 K. How much more luminous is star B compared to star A?

Solution

Equation:

$$L_{\text{A}} = 4\pi R_{\text{A}}{}^2 \sigma T_{\text{A}}{}^4 \text{ and } L_{\text{B}} = 4\pi R_{\text{B}}{}^2 \sigma T_{\text{B}}{}^4$$

Take the ratio of the luminosity of Star A to Star B:
Equation:

$$\frac{L_{\text{B}}}{L_{\text{A}}} = \frac{4\pi R_{\text{B}}{}^2 \sigma T_{\text{B}}{}^4}{4\pi R_{\text{A}}{}^2 \sigma T_{\text{A}}{}^4} = \frac{R_{\text{B}}{}^2 T_{\text{B}}{}^4}{R_{\text{A}}{}^2 T_{\text{A}}{}^4}$$

Because the two stars are the same size, $R_{\text{A}} = R_{\text{B}}$, leaving
Equation:

$$\frac{T_{\text{B}}{}^4}{T_{\text{A}}{}^4} = \frac{(12{,}000\text{ K})^4}{(6000\text{ K})^4} = 2^4 = 16$$

Check Your Learning
Two stars with identical diameters are the same distance away. One has a temperature of 8700 K and the other has a temperature of 2900 K. Which is brighter? How much brighter is it?

Note:
Answer:
The 8700 K star has triple the temperature, so it is $3^4 = 81$ times brighter.

Key Concepts and Summary

The electromagnetic spectrum consists of gamma rays, X-rays, ultraviolet radiation, visible light, infrared, and radio radiation. Many of these wavelengths cannot penetrate the layers of Earth's atmosphere and must be observed from space, whereas others—such as visible light, FM radio and TV—can penetrate to Earth's surface. The emission of electromagnetic

radiation is intimately connected to the temperature of the source. The higher the temperature of an idealized emitter of electromagnetic radiation, the shorter is the wavelength at which the maximum amount of radiation is emitted. The mathematical equation describing this relationship is known as Wien's law: $\lambda_{max} = (3 \times 10^6)/T$. The total power emitted per square meter increases with increasing temperature. The relationship between emitted energy flux and temperature is known as the Stefan-Boltzmann law: $F = \sigma T^4$.

Glossary

blackbody
: an idealized object that absorbs all electromagnetic energy that falls onto it

electromagnetic spectrum
: the whole array or family of electromagnetic waves, from radio to gamma rays

energy flux
: the amount of energy passing through a unit area (for example, 1 square meter) per second; the units of flux are watts per square meter

gamma rays
: photons (of electromagnetic radiation) of energy with wavelengths no longer than 0.01 nanometer; the most energetic form of electromagnetic radiation

infrared
: electromagnetic radiation of wavelength 10^3–10^6 nanometers; longer than the longest (red) wavelengths that can be perceived by the eye, but shorter than radio wavelengths

microwave
: electromagnetic radiation of wavelengths from 1 millimeter to 1 meter; longer than infrared but shorter than radio waves

radio waves

all electromagnetic waves longer than microwaves, including radar waves and AM radio waves

Stefan-Boltzmann law

a formula from which the rate at which a blackbody radiates energy can be computed; the total rate of energy emission from a unit area of a blackbody is proportional to the fourth power of its absolute temperature: $F = \sigma T^4$

ultraviolet

electromagnetic radiation of wavelengths 10 to 400 nanometers; shorter than the shortest visible wavelengths

visible light

electromagnetic radiation with wavelengths of roughly 400–700 nanometers; visible to the human eye

Wien’s law

formula that relates the temperature of a blackbody to the wavelength at which it emits the greatest intensity of radiation

X-rays

electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers; intermediate between those of ultraviolet radiation and gamma rays

Spectroscopy in Astronomy

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the properties of light
- Explain how astronomers learn the composition of a gas by examining its spectral lines
- Discuss the various types of spectra

Electromagnetic radiation carries a lot of information about the nature of stars and other astronomical objects. To extract this information, however, astronomers must be able to study the amounts of energy we receive at different wavelengths of light in fine detail. Let’s examine how we can do this and what we can learn.

Properties of Light

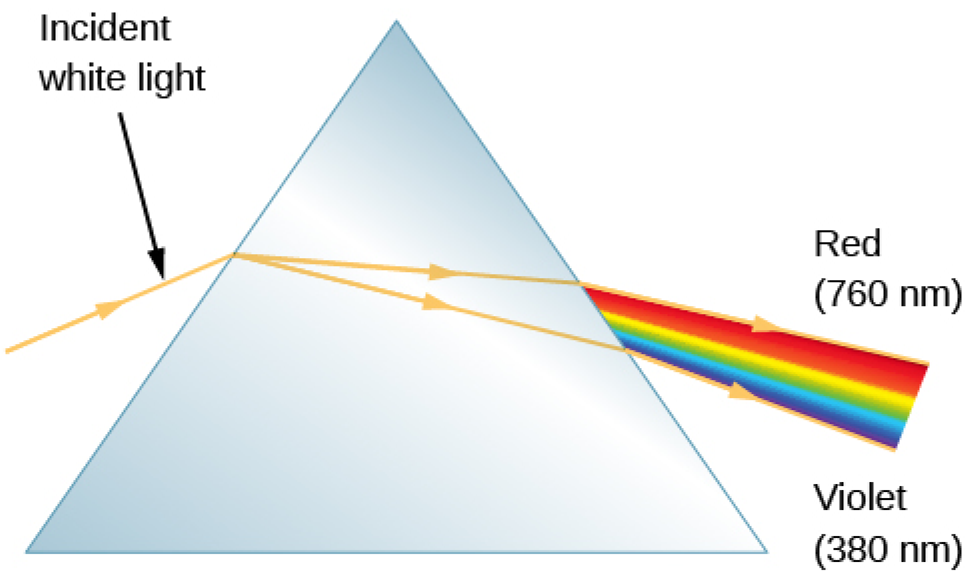
Light exhibits certain behaviors that are important to the design of telescopes and other instruments. For example, light can be *reflected* from a surface. If the surface is smooth and shiny, as with a mirror, the direction of the reflected light beam can be calculated accurately from knowledge of the shape of the reflecting surface. Light is also bent, or *refracted*, when it passes from one kind of transparent material into another—say, from the air into a glass lens.

Reflection and refraction of light are the basic properties that make possible all *optical* instruments (devices that help us to see things better)—from

eyeglasses to giant astronomical telescopes. Such instruments are generally combinations of glass lenses, which bend light according to the principles of refraction, and curved mirrors, which depend on the properties of reflection. Small optical devices, such as eyeglasses or binoculars, generally use lenses, whereas large telescopes depend almost entirely on mirrors for their main optical elements. We will discuss astronomical instruments and their uses more fully in Astronomical Instruments. For now, we turn to another behavior of light, one that is essential for the decoding of light.

In 1672, in the first paper that he submitted to the Royal Society, Sir Isaac Newton described an experiment in which he permitted sunlight to pass through a small hole and then through a prism. Newton found that sunlight, which looks white to us, is actually made up of a mixture of all the colors of the rainbow ([link]).

Action of a Prism.



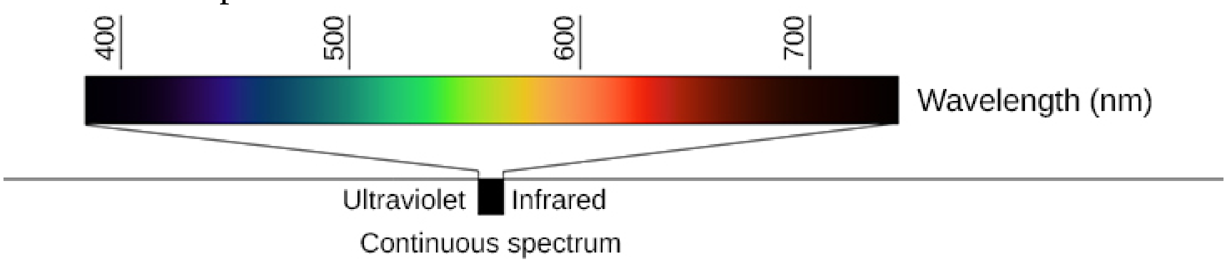
When we pass a beam of white sunlight through a prism, we see a rainbow-colored band of light that we call a continuous spectrum.

[link] shows how light is separated into different colors with a prism—a piece of glass in the shape of a triangle with refracting surfaces. Upon entering one face of the prism, the path of the light is refracted (bent), but

not all of the colors are bent by the same amount. The bending of the beam depends on the wavelength of the light as well as the properties of the material, and as a result, different wavelengths (or colors of light) are bent by different amounts and therefore follow slightly different paths through the prism. The violet light is bent more than the red. This phenomenon is called **dispersion** and explains Newton's rainbow experiment.

Upon leaving the opposite face of the prism, the light is bent again and further dispersed. If the light leaving the prism is focused on a screen, the different wavelengths or colors that make up white light are lined up side by side just like a rainbow ([link]). (In fact, a rainbow is formed by the dispersion of light though raindrops; see The Rainbow feature box.) Because this array of colors is a spectrum of light, the instrument used to disperse the light and form the spectrum is called a **spectrometer**.

Continuous Spectrum.



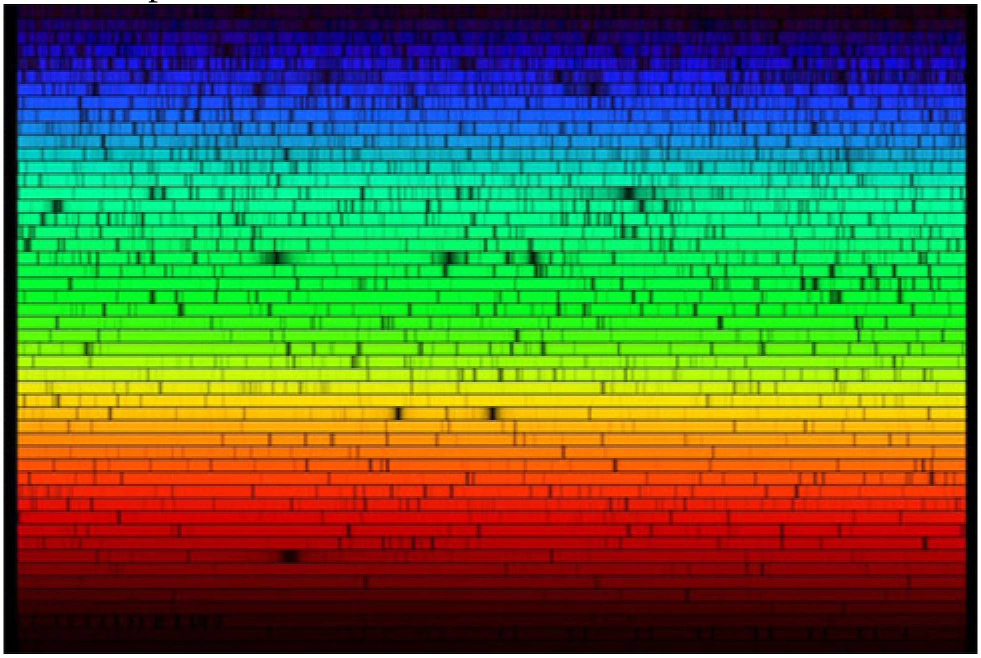
When white light passes through a prism, it is dispersed and forms a continuous spectrum of all the colors. Although it is hard to see in this printed version, in a well-dispersed spectrum, many subtle gradations in color are visible as your eye scans from one end (violet) to the other (red).

The Value of Stellar Spectra

When Newton described the laws of refraction and dispersion in optics, and observed the solar spectrum, all he could see was a continuous band of colors. If the spectrum of the white light from the Sun and stars were simply a continuous rainbow of colors, astronomers would have little interest in the detailed study of a star's spectrum once they had learned its average surface

temperature. In 1802, however, William Wollaston built an improved spectrometer that included a lens to focus the Sun’s spectrum on a screen. With this device, Wollaston saw that the colors were not spread out uniformly, but instead, some ranges of color were missing, appearing as dark bands in the solar spectrum. He mistakenly attributed these lines to natural boundaries between the colors. In 1815, German physicist Joseph Fraunhofer, upon a more careful examination of the solar spectrum, found about 600 such dark lines (missing colors), which led scientists to rule out the boundary hypothesis ([link]).

Visible Spectrum of the Sun.



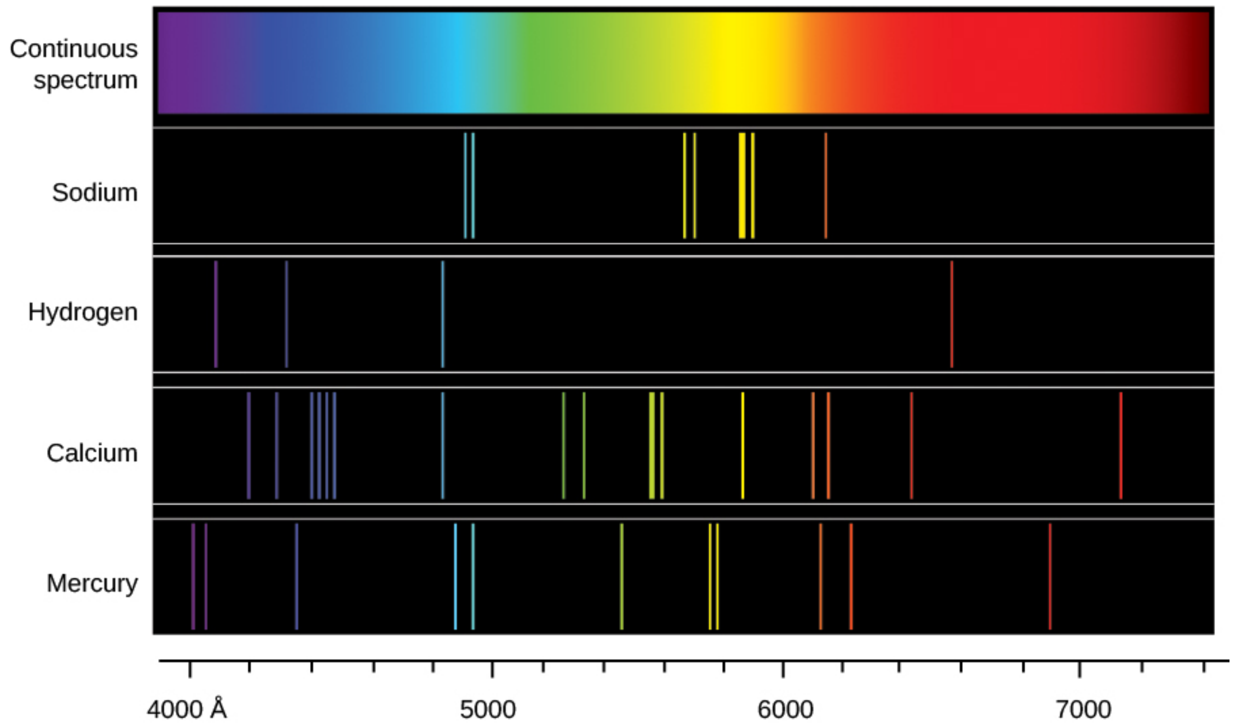
Our star’s spectrum is crossed by dark lines produced by atoms in the solar atmosphere that absorb light at certain wavelengths. (credit: modification of work by Nigel Sharp, NOAO/National Solar Observatory at Kitt Peak/AURA, and the National Science Foundation)

Later, researchers found that similar dark lines could be produced in the spectra (“spectra” is the plural of “spectrum”) of artificial light sources. They did this by passing their light through various apparently transparent substances—usually containers with just a bit of thin gas in them.

These gases turned out not to be transparent at *all* colors: they were quite opaque at a few sharply defined wavelengths. Something in each gas had to be absorbing just a few colors of light and no others. All gases did this, but each different element absorbed a different set of colors and thus showed different dark lines. If the gas in a container consisted of two elements, then light passing through it was missing the colors (showing dark lines) for both of the elements. So it became clear that certain lines in the spectrum “go with” certain elements. This discovery was one of the most important steps forward in the history of astronomy.

What would happen if there were no continuous spectrum for our gases to remove light from? What if, instead, we heated the same thin gases until they were hot enough to glow with their own light? When the gases were heated, a spectrometer revealed no continuous spectrum, but several separate bright lines. That is, these hot gases emitted light only at certain specific wavelengths or colors.

When the gas was pure hydrogen, it would emit one pattern of colors; when it was pure sodium, it would emit a different pattern. A mixture of hydrogen and sodium emitted both sets of spectral lines. The colors the gases emitted when they were heated were the very same colors as those they had absorbed when a continuous source of light was behind them. From such experiments, scientists began to see that different substances showed distinctive *spectral signatures* by which their presence could be detected ([link]). Just as your signature allows the bank to identify you, the unique pattern of colors for each type of atom (its spectrum) can help us identify which element or elements are in a gas.
Continuous Spectrum and Line Spectra from Different Elements.



Each type of glowing gas (each element) produces its own unique pattern of lines, so the composition of a gas can be identified by its spectrum. The spectra of sodium, hydrogen, calcium, and mercury gases are shown here.

Types of Spectra

In these experiments, then, there were three different types of spectra. A **continuous spectrum** (formed when a solid or very dense gas gives off radiation) is an array of all wavelengths or colors of the rainbow. A continuous spectrum can serve as a backdrop from which the atoms of much less dense gas can absorb light. A dark line, or **absorption spectrum**, consists of a series or pattern of dark lines—missing colors—superimposed upon the continuous spectrum of a source. A bright line, or **emission spectrum**, appears as a pattern or series of bright lines; it consists of light in which only certain discrete wavelengths are present. ([link] shows an absorption spectrum, whereas [link] shows the emission spectrum of a

number of common elements along with an example of a continuous spectrum.)

When we have a hot, thin gas, each particular chemical element or compound produces its own characteristic pattern of spectral lines—its spectral signature. No two types of atoms or molecules give the same patterns. In other words, each particular gas can absorb or emit only certain wavelengths of the light peculiar to that gas. In contrast, absorption spectra occur when passing white light through a cool, thin gas. The temperature and other conditions determine whether the lines are bright or dark (whether light is absorbed or emitted), but the wavelengths of the lines for any element are the same in either case. It is the precise pattern of wavelengths that makes the signature of each element unique. Liquids and solids can also generate spectral lines or bands, but they are broader and less well defined—and hence, more difficult to interpret. Spectral analysis, however, can be quite useful. It can, for example, be applied to light reflected off the surface of a nearby asteroid as well as to light from a distant galaxy.

The dark lines in the solar spectrum thus give evidence of certain chemical elements between us and the Sun absorbing those wavelengths of sunlight. Because the space between us and the Sun is pretty empty, astronomers realized that the atoms doing the absorbing must be in a thin atmosphere of cooler gas around the Sun. This outer atmosphere is not all that different from the rest of the Sun, just thinner and cooler. Thus, we can use what we learn about its composition as an indicator of what the whole Sun is made of. Similarly, we can use the presence of absorption and emission lines to analyze the composition of other stars and clouds of gas in space.

Such analysis of spectra is the key to modern astronomy. Only in this way can we “sample” the stars, which are too far away for us to visit. Encoded in the electromagnetic radiation from celestial objects is clear information about the chemical makeup of these objects. Only by understanding what the stars were made of could astronomers begin to form theories about what made them shine and how they evolved.

In 1860, German physicist Gustav Kirchhoff became the first person to use spectroscopy to identify an element in the Sun when he found the spectral signature of sodium gas. In the years that followed, astronomers found

many other chemical elements in the Sun and stars. In fact, the element helium was found first in the Sun from its spectrum and only later identified on Earth. (The word “helium” comes from *helios*, the Greek name for the Sun.)

Why are there specific lines for each element? The answer to that question was not found until the twentieth century; it required the development of a model for the atom. We therefore turn next to a closer examination of the atoms that make up all matter.

Note:
The Rainbow
Rainbows are an excellent illustration of the dispersion of sunlight. You have a good chance of seeing a rainbow any time you are between the Sun and a rain shower, as illustrated in [link]. The raindrops act like little prisms and break white light into the spectrum of colors. Suppose a ray of sunlight encounters a raindrop and passes into it. The light changes direction—is refracted—when it passes from air to water; the blue and violet light are refracted more than the red. Some of the light is then reflected at the backside of the drop and reemerges from the front, where it is again refracted. As a result, the white light is spread out into a rainbow of colors.
Rainbow Refraction.

(a) This diagram shows how light from the Sun, which is located behind the observer, can be refracted by raindrops to produce (b) a rainbow. (c) Refraction separates white light into its component colors.

Note that violet light lies above the red light after it emerges from the raindrop. When you look at a rainbow, however, the red light is higher in the sky. Why? Look again at [link]. If the observer looks at a raindrop that is high in the sky, the violet light passes over her head and the red light enters her eye. Similarly, if the observer looks at a raindrop that is low in the sky, the violet light reaches her eye and the drop appears violet, whereas the red light from that same drop strikes the ground and is not seen. Colors of intermediate wavelengths are refracted to the eye by drops that are intermediate in altitude between the drops that appear violet and the ones that appear red. Thus, a single rainbow always has red on the outside and violet on the inside.

Key Concepts and Summary

A spectrometer is a device that forms a spectrum, often utilizing the phenomenon of dispersion. The light from an astronomical source can consist of a continuous spectrum, an emission (bright line) spectrum, or an absorption (dark line) spectrum. Because each element leaves its spectral signature in the pattern of lines we observe, spectral analyses reveal the composition of the Sun and stars.

Glossary

absorption spectrum
: a series or pattern of dark lines superimposed on a continuous spectrum

continuous spectrum
: a spectrum of light composed of radiation of a continuous range of wavelengths or colors, rather than only certain discrete wavelengths

dispersion
: separation of different wavelengths of white light through refraction of different amounts

emission spectrum

a series or pattern of bright lines superimposed on a continuous spectrum

spectrometer

an instrument for obtaining a spectrum; in astronomy, usually attached to a telescope to record the spectrum of a star, galaxy, or other astronomical object

The Structure of the Atom

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the structure of atoms and the components of nuclei
- Explain the behavior of electrons within atoms and how electrons interact with light to move among energy levels

The idea that matter is composed of tiny particles called atoms is at least 25 centuries old. It took until the twentieth century, however, for scientists to invent instruments that permitted them to probe inside an atom and find that it is not, as had been thought, hard and indivisible. Instead, the atom is a complex structure composed of still smaller particles.

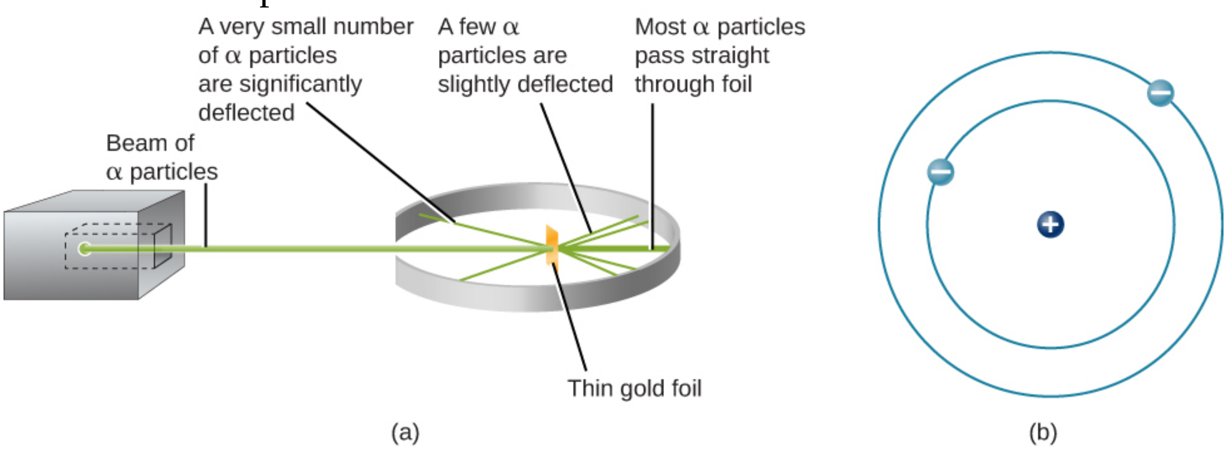
Probing the Atom

The first of these smaller particles was discovered by British physicist James (J. J.) Thomson in 1897. Named the *electron*, this particle is negatively charged. (It is the flow of these particles that produces currents of electricity, whether in lightning bolts or in the wires leading to your lamp.) Because an atom in its normal state is electrically neutral, each electron in an atom must be balanced by the same amount of positive charge.

The next step was to determine where in the atom the positive and negative charges are located. In 1911, British physicist Ernest Rutherford devised an experiment that provided part of the answer to this question. He bombarded

an extremely thin piece of gold foil, only about 400 atoms thick, with a beam of alpha particles ([link]). *Alpha particles* (α particles) are helium atoms that have lost their electrons and thus are positively charged. Most of these particles passed though the gold foil just as if it and the atoms in it were nearly empty space. About 1 in 8000 of the alpha particles, however, completely reversed direction and bounced backward from the foil. Rutherford wrote, “It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.”

Rutherford’s Experiment.



(a) When Rutherford allowed α particles from a radioactive source to strike a target of gold foil, he found that, although most of them went straight through, some rebounded back in the direction from which they came. (b) From this experiment, he concluded that the atom must be constructed like a miniature solar system, with the positive charge concentrated in the nucleus and the negative charge orbiting in the large volume around the nucleus. Note that this drawing is not to scale; the electron orbits are much larger relative to the size of the nucleus.

The only way to account for the particles that reversed direction when they hit the gold foil was to assume that nearly all of the mass, as well as all of the positive charge in each individual gold atom, is concentrated in a tiny center or **nucleus**. When a positively charged alpha particle strikes a nucleus, it reverses direction, much as a cue ball reverses direction when it

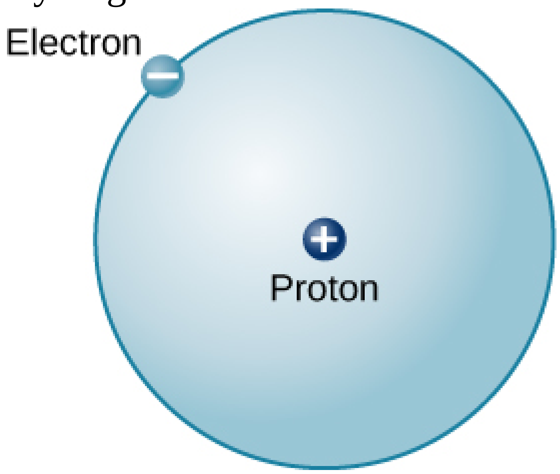
strikes another billiard ball. Rutherford’s model placed the other type of charge—the negative electrons—in orbit around this nucleus.

Rutherford’s model required that the electrons be in motion. Positive and negative charges attract each other, so stationary electrons would fall into the positive nucleus. Also, because both the electrons and the nucleus are extremely small, most of the atom is empty, which is why nearly all of Rutherford’s particles were able to pass right through the gold foil without colliding with anything. Rutherford’s model was a very successful explanation of the experiments he conducted, although eventually scientists would discover that even the nucleus itself has structure.

The Atomic Nucleus

The simplest possible atom (and the most common one in the Sun and stars) is hydrogen. The nucleus of ordinary hydrogen contains a single proton. Moving around this proton is a single electron. The mass of an electron is nearly 2000 times smaller than the mass of a proton; the electron carries an amount of charge exactly equal to that of the proton but opposite in sign ([link]). Opposite charges attract each other, so it is an electromagnetic force that holds the proton and electron together, just as gravity is the force that keeps planets in orbit around the Sun.

Hydrogen Atom.

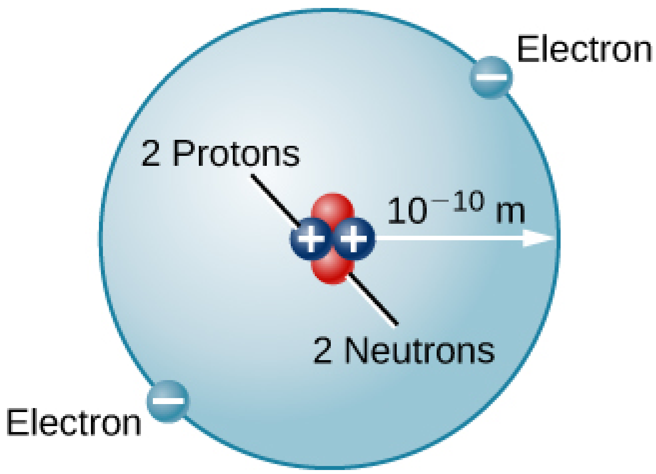


This is a schematic diagram of a hydrogen atom in its lowest energy state, also

called the ground state. The proton and electron have equal but opposite charges, which exert an electromagnetic force that binds the hydrogen atom together. In the illustration, the size of the particles is exaggerated so that you can see them; they are not to scale. They are also shown much closer than they would actually be as it would take more than an entire page to show their actual distance to scale.

There are many other types of atoms in nature. Helium, for example, is the second-most abundant element in the Sun. Helium has two protons in its nucleus instead of the single proton that characterizes hydrogen. In addition, the helium nucleus contains two neutrons, particles with a mass comparable to that of the proton but with no electric charge. Moving around this nucleus are two electrons, so the total net charge of the helium atom is also zero ([link]).

Helium Atom.

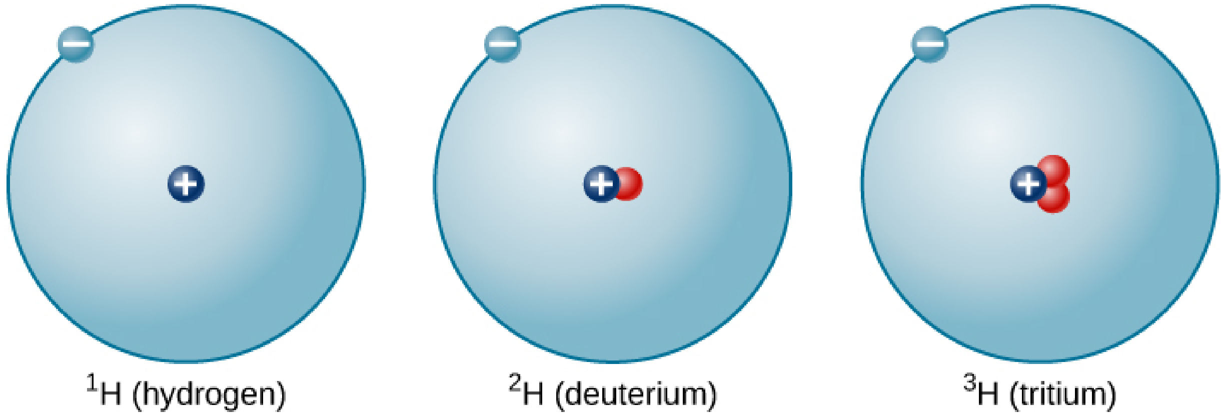


Here we see a schematic diagram of a helium atom in its lowest energy state. Two protons are present in the nucleus of all helium atoms. In the most common variety of helium, the nucleus also contains two neutrons, which have nearly the same mass as the proton but carry no charge. Two electrons orbit the nucleus.

From this description of hydrogen and helium, perhaps you have guessed the pattern for building up all the elements (different types of atoms) that we find in the universe. The type of element is determined by the number of protons in the nucleus of the atom. For example, any atom with six protons is the element carbon, with eight protons is oxygen, with 26 is iron, and with 92 is uranium. On Earth, a typical atom has the same number of electrons as protons, and these electrons follow complex orbital patterns around the nucleus. Deep inside stars, however, it is so hot that the electrons get loose from the nucleus and (as we shall see) lead separate yet productive lives.

The ratio of neutrons to protons increases as the number of protons increases, but each element is unique. The number of neutrons is not necessarily the same for all atoms of a given element. For example, most hydrogen atoms contain no neutrons at all. There are, however, hydrogen atoms that contain one proton and one neutron, and others that contain one proton and two neutrons. The various types of hydrogen nuclei with different numbers of neutrons are called **isotopes** of hydrogen ([link]), and all other elements have isotopes as well. You can think of isotopes as siblings in the same element “family”—closely related but with different characteristics and behaviors.

Isotopes of Hydrogen.



A single proton in the nucleus defines the atom to be hydrogen, but there may be zero, one, or two neutrons. The most common isotope of hydrogen is the one with only a single proton and no neutrons.

Note:
To explore the structure of atoms, go to the PhET Build an Atom website where you can add protons, neutrons, or electrons to a model and the name of the element you have created will appear. You can also see the net charge, the mass number, whether it is stable or unstable, and whether it is an ion or a neutral atom.

The Bohr Atom

Rutherford’s model for atoms has one serious problem. Maxwell’s theory of electromagnetic radiation says that when electrons change either speed or the direction of motion, they must emit energy. Orbiting electrons constantly change their direction of motion, so they should emit a constant stream of energy. Applying Maxwell’s theory to Rutherford’s model, all electrons should spiral into the nucleus of the atom as they lose energy, and this collapse should happen very quickly—in about 10^{-16} seconds.

It was Danish physicist Niels Bohr (1885–1962) who solved the mystery of how electrons remain in orbit. He was trying to develop a model of the atom that would also explain certain regularities observed in the spectrum of hydrogen. He suggested that the spectrum of hydrogen can be understood if we assume that orbits of only certain sizes are possible for the electron. Bohr further assumed that as long as the electron moves in only one of these allowed orbits, it radiates no energy: its energy would change only if it moved from one orbit to another.

This suggestion, in the words of science historian Abraham Pais, was “one of the most audacious hypotheses ever introduced in physics.” If something equivalent were at work in the everyday world, you might find that, as you went for a walk after astronomy class, nature permitted you to walk two steps per minute, five steps per minute, and 12 steps per minute, but no speeds in between. No matter how you tried to move your legs, only certain walking speeds would be permitted. To make things more bizarre, it would take no effort to walk at any one of the allowed speeds, but it would be difficult to change from one speed to another. Luckily, no such rules apply at the level of human behavior. But at the microscopic level of the atom, experiment after experiment has confirmed the validity of Bohr’s strange idea. Bohr’s suggestions became one of the foundations of the new (and much more sophisticated) model of the subatomic world called quantum mechanics.

In Bohr’s model, if the electron moves from one orbit to another closer to the atomic nucleus, it must give up some energy in the form of electromagnetic radiation. If the electron goes from an inner orbit to one farther from the nucleus, however, it requires some additional energy. One way to obtain the necessary energy is to absorb electromagnetic radiation that may be streaming past the atom from an outside source.

A key feature of Bohr’s model is that each of the permitted electron orbits around a given atom has a certain energy value; we therefore can think of each orbit as an **energy level**. To move from one orbit to another (which will have its own specific energy value) requires a change in the electron’s energy—a change determined by the difference between the two energy values. If the electron goes to a lower level, the energy difference will be

given off; if the electron goes to a higher level, the energy difference must be obtained from somewhere else. Each jump (or transition) to a different level has a fixed and definite energy change associated with it.

A crude analogy for this situation might be life in a tower of luxury apartments where the rent is determined by the quality of the view. Such a building has certain, definite numbered levels or floors on which apartments are located. No one can live on floor 5.37 or 22.5. In addition, the rent gets higher as you go up to higher floors. If you want to exchange an apartment on the twentieth floor for one on the second floor, you will not owe as much rent. However, if you want to move from the third floor to the twenty-fifth floor, your rent will increase. In an atom, too, the “cheapest” place for an electron to live is the lowest possible level, and energy is required to move to a higher level.

Here we have one of the situations where it is easier to think of electromagnetic radiation as particles (photons) rather than as waves. As electrons move from one level to another, they give off or absorb little packets of energy. When an electron moves to a higher level, it absorbs a photon of just the right energy (provided one is available). When it moves to a lower level, it emits a photon with the exact amount of energy it no longer needs in its “lower-cost living situation.”

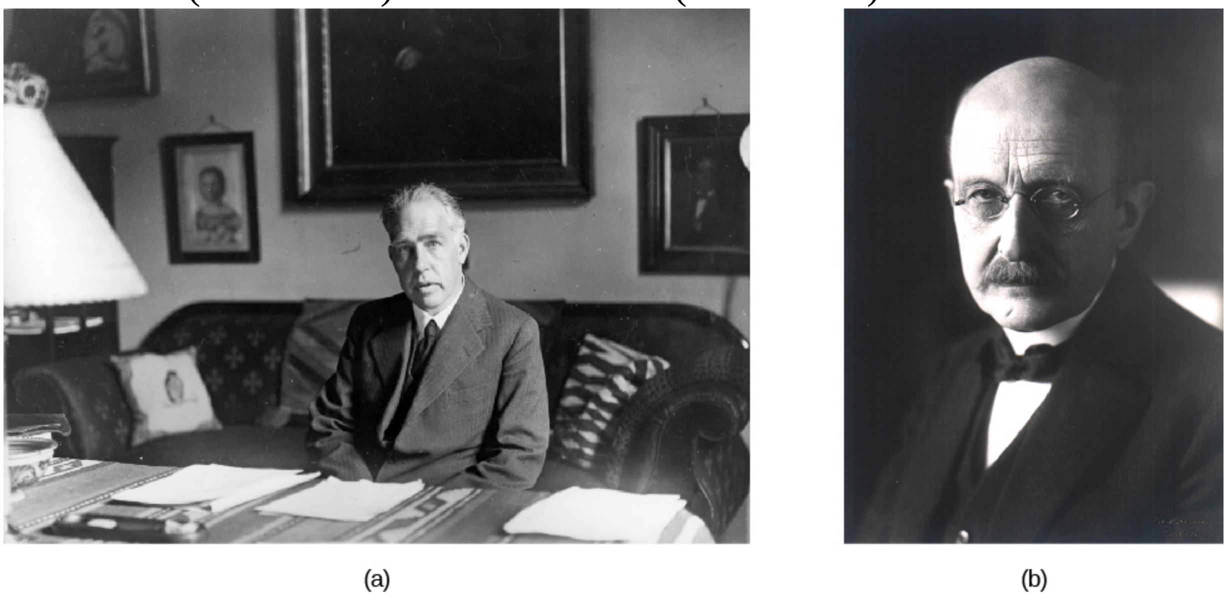
The photon and wave perspectives must be equivalent: light is light, no matter how we look at it. Thus, each photon carries a certain amount of energy that is proportional to the frequency (f) of the wave it represents. The value of its energy (E) is given by the formula
Equation:

$$E = hf$$

where the constant of proportionality, h, is called Planck’s constant.

The constant is named for Max Planck, the German physicist who was one of the originators of the quantum theory ([link]). If metric units are used (that is, if energy is measured in joules and frequency in hertz), then Planck’s constant has the value $h = 6.626 \times 10^{-34}$ joule-seconds (J-s).

Higher-energy photons correspond to higher-frequency waves (which have a shorter wavelength); lower-energy photons are waves of lower frequency. Niels Bohr (1885–1962) and Max Planck (1858–1947).



(a) Bohr, shown at his desk in this 1935 photograph, and (b) Planck helped us understand the energy behavior of photons.

To take a specific example, consider a calcium atom inside the Sun's atmosphere in which an electron jumps from a lower level to a higher level. To do this, it needs about 5×10^{-19} joules of energy, which it can conveniently obtain by absorbing a passing photon of that energy coming from deeper inside the Sun. This photon is equivalent to a wave of light whose frequency is about 7.5×10^{14} hertz and whose wavelength is about 3.9×10^{-7} meters (393 nanometers), in the deep violet part of the visible light spectrum. Although it may seem strange at first to switch from picturing light as a photon (or energy packet) to picturing it as a wave, such switching has become second nature to astronomers and can be a handy tool for doing calculations about spectra.

Example:
The Energy of a Photon

Now that we know how to calculate the wavelength and frequency of a photon, we can use this information, along with Planck's constant, to determine how much energy each photon carries. How much energy does a red photon of wavelength 630 nm have?

Solution

First, as we learned earlier, we can find the frequency of the photon:

Equation:

$$f = \frac{c}{\lambda} = \frac{3 \times 10^{8}\,\text{m/s}}{630 \times 10^{-9}\,\text{m}} = 4.8 \times 10^{14}\,\text{Hz}$$

Next, we can use Planck's constant to determine the energy (remember that a Hz is the same as 1/s):

Equation:

$$E = hf = \left(6.626 \times 10^{-34}\,\text{J-s}\right)\left(4.8 \times 10^{14}\,\text{Hz}(1/\text{s})\right) = 3.2 \times 10^{-19}\,\text{J}$$

Check Your Learning

What is the energy of a yellow photon with a frequency of 5.5×10^{14} Hz?

Note:

Answer:

$E = hf = \left(6.626 \times 10^{-34}\text{J-s}\right)\left(5.5 \times 10^{14}\text{Hz}\right) = 3.6 \times 10^{-19}\,\text{J}$

Key Concepts and Summary

Atoms consist of a nucleus containing one or more positively charged protons. All atoms except hydrogen can also contain one or more neutrons in the nucleus. Negatively charged electrons orbit the nucleus. The number of protons defines an element (hydrogen has one proton, helium has two, and so on) of the atom. Nuclei with the same number of protons but

different numbers of neutrons are different isotopes of the same element. In the Bohr model of the atom, electrons on permitted orbits (or energy levels) don’t give off any electromagnetic radiation. But when electrons go from lower levels to higher ones, they must absorb a photon of just the right energy, and when they go from higher levels to lower ones, they give off a photon of just the right energy. The energy of a photon is connected to the frequency of the electromagnetic wave it represents by Planck’s formula, $E = hf$.

Glossary

energy level
: a particular level, or amount, of energy possessed by an atom or ion above the energy it possesses in its least energetic state; also used to refer to the states of energy an electron can have in an atom

isotope
: any of two or more forms of the same element whose atoms have the same number of protons but different numbers of neutrons

nucleus (of an atom)
: the massive part of an atom, composed mostly of protons and neutrons, and about which the electrons revolve

Formation of Spectral Lines

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how emission line spectra and absorption line spectra are formed
- Describe what ions are and how they are formed
- Explain how spectral lines and ionization levels in a gas can help us determine its temperature

We can use Bohr’s model of the atom to understand how spectral lines are formed. The concept of energy levels for the electron orbits in an atom leads naturally to an explanation of why atoms absorb or emit only specific energies or wavelengths of light.

The Hydrogen Spectrum

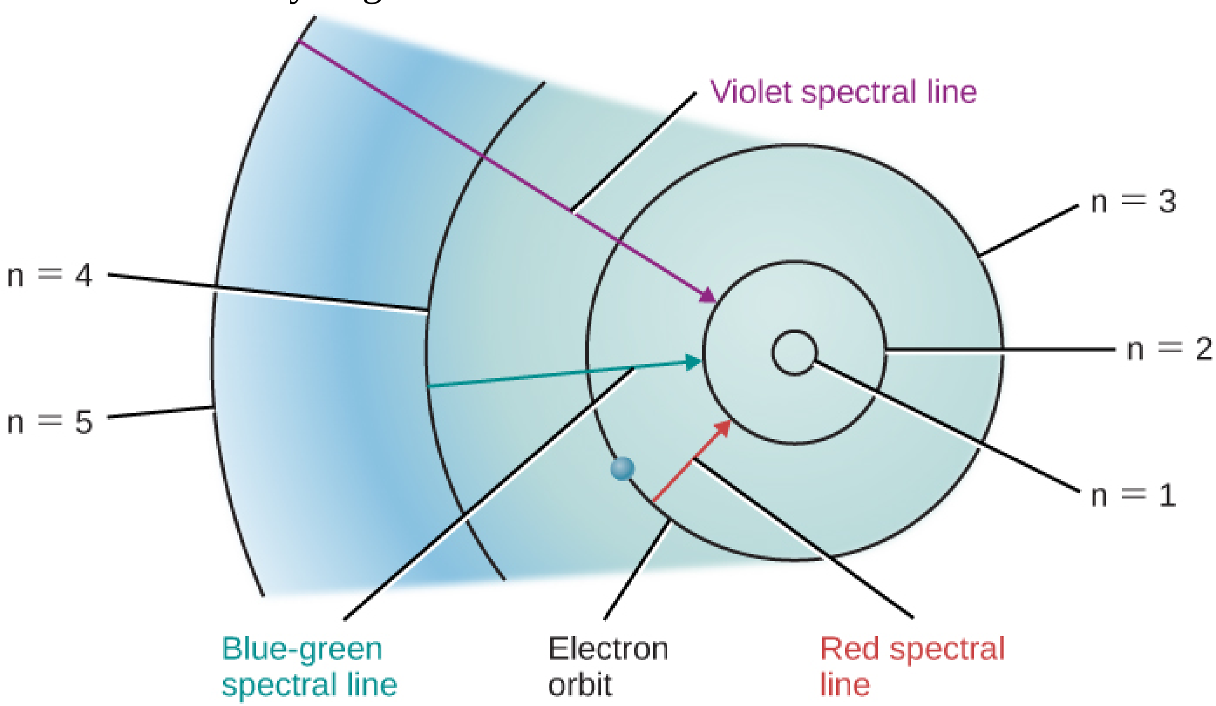
Let’s look at the hydrogen atom from the perspective of the Bohr model. Suppose a beam of white light (which consists of photons of all visible wavelengths) shines through a gas of atomic hydrogen. A photon of wavelength 656 nanometers has just the right energy to raise an electron in a hydrogen atom from the second to the third orbit. Thus, as all the photons of different energies (or wavelengths or colors) stream by the hydrogen atoms, photons with *this* particular wavelength can be absorbed by those atoms whose electrons are orbiting on the second level. When they are absorbed, the electrons on the second level will move to the third level, and

a number of the photons of this wavelength and energy will be missing from the general stream of white light.

Other photons will have the right energies to raise electrons from the second to the fourth orbit, or from the first to the fifth orbit, and so on. Only photons with these exact energies can be absorbed. All of the other photons will stream past the atoms untouched. Thus, hydrogen atoms absorb light at only certain wavelengths and produce dark lines at those wavelengths in the spectrum we see.

Suppose we have a container of hydrogen gas through which a whole series of photons is passing, allowing many electrons to move up to higher levels. When we turn off the light source, these electrons “fall” back down from larger to smaller orbits and emit photons of light—but, again, only light of those energies or wavelengths that correspond to the energy difference between permissible orbits. The orbital changes of hydrogen electrons that give rise to some spectral lines are shown in [link].

Bohr Model for Hydrogen.



In this simplified model of a hydrogen atom, the concentric circles shown represent permitted orbits or energy levels. An electron in a hydrogen atom can only exist in one of these energy levels (or states).

The closer the electron is to the nucleus, the more tightly bound the electron is to the nucleus. By absorbing energy, the electron can move to energy levels farther from the nucleus (and even escape if enough energy is absorbed).

Similar pictures can be drawn for atoms other than hydrogen. However, because these other atoms ordinarily have more than one electron each, the orbits of their electrons are much more complicated, and the spectra are more complex as well. For our purposes, the key conclusion is this: *each type of atom has its own unique pattern of electron orbits, and no two sets of orbits are exactly alike*. This means that each type of atom shows its own unique set of spectral lines, produced by electrons moving between its unique set of orbits.

Astronomers and physicists have worked hard to learn the lines that go with each element by studying the way atoms absorb and emit light in laboratories here on Earth. Then they can use this knowledge to identify the elements in celestial bodies. In this way, we now know the chemical makeup of not just any star, but even galaxies of stars so distant that their light started on its way to us long before Earth had even formed.

Energy Levels and Excitation

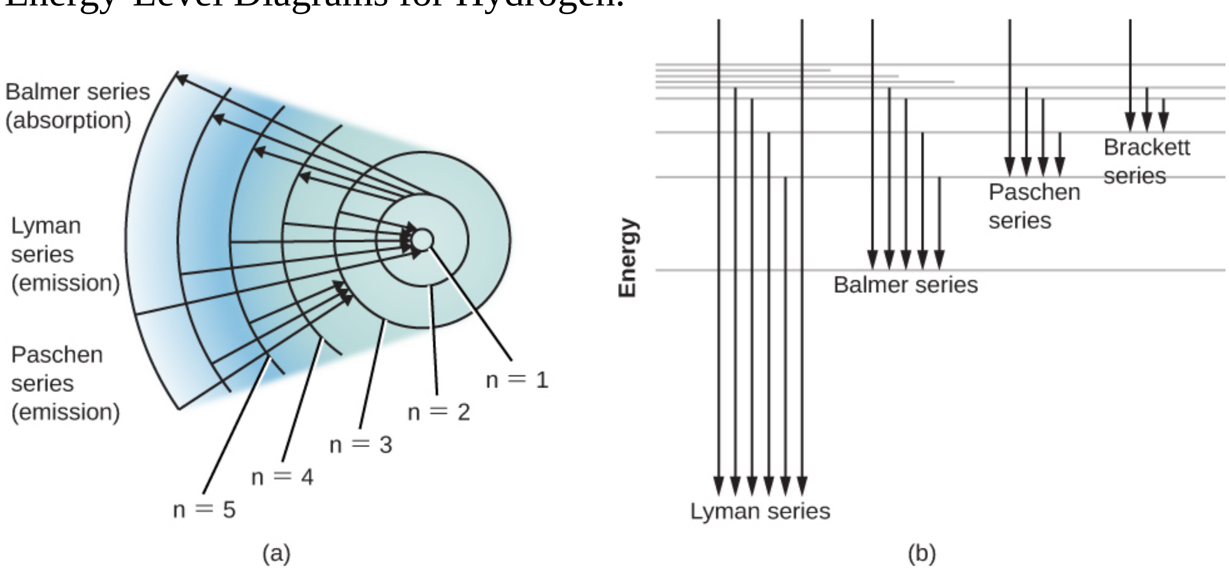
Bohr's model of the hydrogen atom was a great step forward in our understanding of the atom. However, we know today that atoms cannot be represented by quite so simple a picture. For example, the concept of sharply defined electron orbits is not really correct; however, at the level of this introductory course, the notion that only certain discrete energies are allowable for an atom is very useful. The energy levels we have been discussing can be thought of as representing certain average distances of the electron's possible orbits from the atomic nucleus.

Ordinarily, an atom is in the state of lowest possible energy, its **ground state**. In the Bohr model of the hydrogen atom, the ground state corresponds to the electron being in the innermost orbit. An atom can

absorb energy, which raises it to a higher energy level (corresponding, in the simple Bohr picture, to an electron's movement to a larger orbit)—this is referred to as **excitation**. The atom is then said to be in an *excited state*. Generally, an atom remains excited for only a very brief time. After a short interval, typically a hundred-millionth of a second or so, it drops back spontaneously to its ground state, with the simultaneous emission of light. The atom may return to its lowest state in one jump, or it may make the transition in steps of two or more jumps, stopping at intermediate levels on the way down. With each jump, it emits a photon of the wavelength that corresponds to the energy difference between the levels at the beginning and end of that jump.

An energy-level diagram for a hydrogen atom and several possible atomic transitions are shown in [link]. When we measure the energies involved as the atom jumps between levels, we find that the transitions to or from the ground state, called the *Lyman series* of lines, result in the emission or absorption of ultraviolet photons. But the transitions to or from the first excited state (labeled n = 2 in part (a) of [link]), called the Balmer series, produce emission or absorption in visible light. In fact, it was to explain this Balmer series that Bohr first suggested his model of the atom.

Energy-Level Diagrams for Hydrogen.



(a) Here we follow the emission or absorption of photons by a hydrogen atom according to the Bohr model. Several different series of spectral lines are shown, corresponding to transitions of electrons from

or to certain allowed orbits. Each series of lines that terminates on a specific inner orbit is named for the physicist who studied it. At the top, for example, you see the Balmer series, and arrows show electrons jumping from the second orbit (n = 2) to the third, fourth, fifth, and sixth orbits. Each time a “poor” electron from a lower level wants to rise to a higher position in life, it must absorb energy to do so. It can absorb the energy it needs from passing waves (or photons) of light. The next set of arrows (Lyman series) show electrons falling down to the first orbit from different (higher) levels. Each time a “rich” electron goes downward toward the nucleus, it can afford to give off (emit) some energy it no longer needs. (b) At higher and higher energy levels, the levels become more and more crowded together, approaching a limit. The region above the top line represents energies at which the atom is ionized (the electron is no longer attached to the atom). Each series of arrows represents electrons falling from higher levels to lower ones, releasing photons or waves of energy in the process.

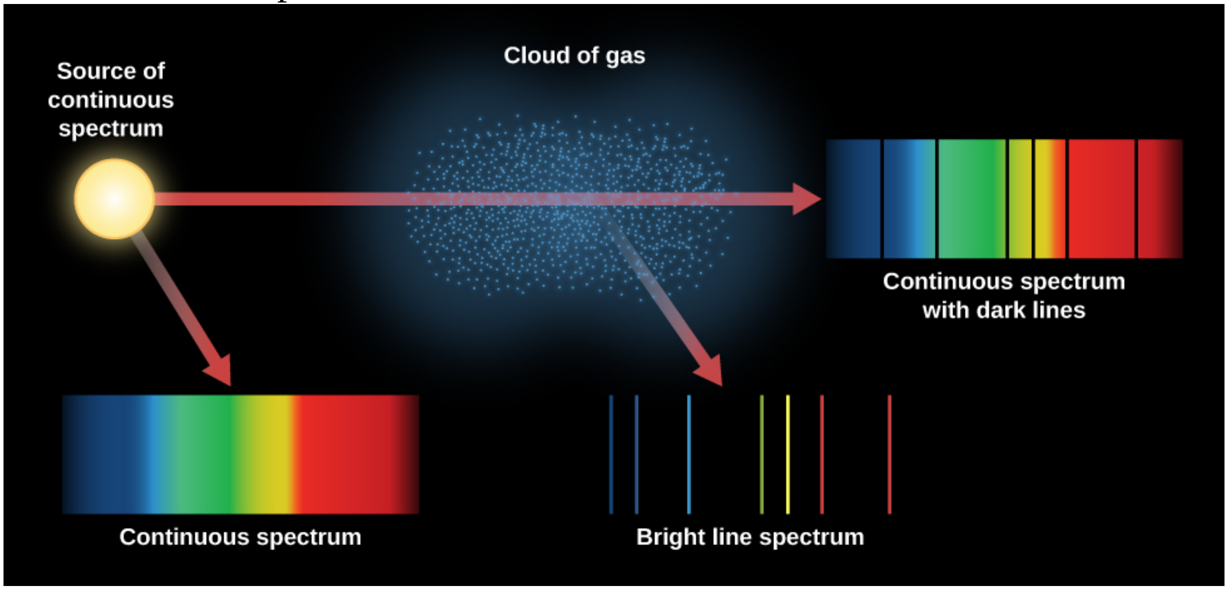
Atoms that have absorbed specific photons from a passing beam of white light and have thus become excited generally de-excite themselves and emit that light again in a very short time. You might wonder, then, why *dark* spectral lines are ever produced. In other words, why doesn’t this reemitted light quickly “fill in” the darker absorption lines?

Imagine a beam of white light coming toward you through some cooler gas. Some of the reemitted light *is* actually returned to the beam of white light you see, but this fills in the absorption lines only to a slight extent. The reason is that the atoms in the gas reemit light *in all directions*, and only a small fraction of the reemitted light is in the direction of the original beam (toward you). In a star, much of the reemitted light actually goes in directions leading back into the star, which does observers outside the star no good whatsoever.

[link] summarizes the different kinds of spectra we have discussed. An incandescent lightbulb produces a continuous spectrum. When that

continuous spectrum is viewed through a thinner cloud of gas, an absorption line spectrum can be seen superimposed on the continuous spectrum. If we look only at a cloud of excited gas atoms (with no continuous source seen behind it), we see that the excited atoms give off an emission line spectrum.

Three Kinds of Spectra.



When we see a lightbulb or other source of continuous radiation, all the colors are present. When the continuous spectrum is seen through a thinner gas cloud, the cloud's atoms produce absorption lines in the continuous spectrum. When the excited cloud is seen without the continuous source behind it, its atoms produce emission lines. We can learn which types of atoms are in the gas cloud from the pattern of absorption or emission lines.

Atoms in a hot gas are moving at high speeds and continually colliding with one another and with any loose electrons. They can be excited (electrons moving to a higher level) and de-excited (electrons moving to a lower level) by these collisions as well as by absorbing and emitting light. The speed of atoms in a gas depends on the temperature. When the temperature is higher, so are the speed and energy of the collisions. The hotter the gas, therefore, the more likely that electrons will occupy the outermost orbits, which

correspond to the highest energy levels. This means that the level where electrons *start* their upward jumps in a gas can serve as an indicator of how hot that gas is. In this way, the absorption lines in a spectrum give astronomers information about the temperature of the regions where the lines originate.

Note:
Use this simulation to play with a hydrogen atom and see what happens when electrons move to higher levels and then give off photons as they go to a lower level.

Ionization

We have described how certain discrete amounts of energy can be absorbed by an atom, raising it to an excited state and moving one of its electrons farther from its nucleus. If enough energy is absorbed, the electron can be completely removed from the atom—this is called **ionization**. The atom is then said to be ionized. The minimum amount of energy required to remove one electron from an atom in its ground state is called its ionization energy.

Still-greater amounts of energy must be absorbed by the now-ionized atom (called an **ion**) to remove an additional electron deeper in the structure of the atom. Successively greater energies are needed to remove the third, fourth, fifth—and so on—electrons from the atom. If enough energy is available, an atom can become completely ionized, losing all of its electrons. A hydrogen atom, having only one electron to lose, can be ionized only once; a helium atom can be ionized twice; and an oxygen atom up to eight times. When we examine regions of the cosmos where there is a great deal of energetic radiation, such as the neighborhoods where hot young stars have recently formed, we see a lot of ionization going on.

An atom that has become positively ionized has lost a negative charge—the missing electron—and thus is left with a net positive charge. It therefore exerts a strong attraction on any free electron. Eventually, one or more

electrons will be captured and the atom will become neutral (or ionized to one less degree) again. During the electron-capture process, the atom emits one or more photons. Which photons are emitted depends on whether the electron is captured at once to the lowest energy level of the atom or stops at one or more intermediate levels on its way to the lowest available level.

Just as the excitation of an atom can result from a collision with another atom, ion, or electron (collisions with electrons are usually most important), so can ionization. The rate at which such collisional ionizations occur depends on the speeds of the atoms and hence on the temperature of the gas—the hotter the gas, the more of its atoms will be ionized.

The rate at which ions and electrons recombine also depends on their relative speeds—that is, on the temperature. In addition, it depends on the density of the gas: the higher the density, the greater the chance for recapture, because the different kinds of particles are crowded more closely together. From a knowledge of the temperature and density of a gas, it is possible to calculate the fraction of atoms that have been ionized once, twice, and so on. In the Sun, for example, we find that most of the hydrogen and helium atoms in its atmosphere are neutral, whereas most of the calcium atoms, as well as many other heavier atoms, are ionized once.

The energy levels of an ionized atom are entirely different from those of the same atom when it is neutral. Each time an electron is removed from the atom, the energy levels of the ion, and thus the wavelengths of the spectral lines it can produce, change. This helps astronomers differentiate the ions of a given element. Ionized hydrogen, having no electron, can produce no absorption lines.

Key Concepts and Summary

When electrons move from a higher energy level to a lower one, photons are emitted, and an emission line can be seen in the spectrum. Absorption lines are seen when electrons absorb photons and move to higher energy levels. Since each atom has its own characteristic set of energy levels, each is associated with a unique pattern of spectral lines. This allows astronomers to determine what elements are present in the stars and in the

clouds of gas and dust among the stars. An atom in its lowest energy level is in the ground state. If an electron is in an orbit other than the least energetic one possible, the atom is said to be excited. If an atom has lost one or more electrons, it is called an ion and is said to be ionized. The spectra of different ions look different and can tell astronomers about the temperatures of the sources they are observing.

Glossary

excitation
: the process of giving an atom or an ion an amount of energy greater than it has in its lowest energy (ground) state

ground state
: the lowest energy state of an atom

ion
: an atom that has become electrically charged by the addition or loss of one or more electrons

ionization
: the process by which an atom gains or loses electrons

The Doppler Effect

LEARNING OBJECTIVES

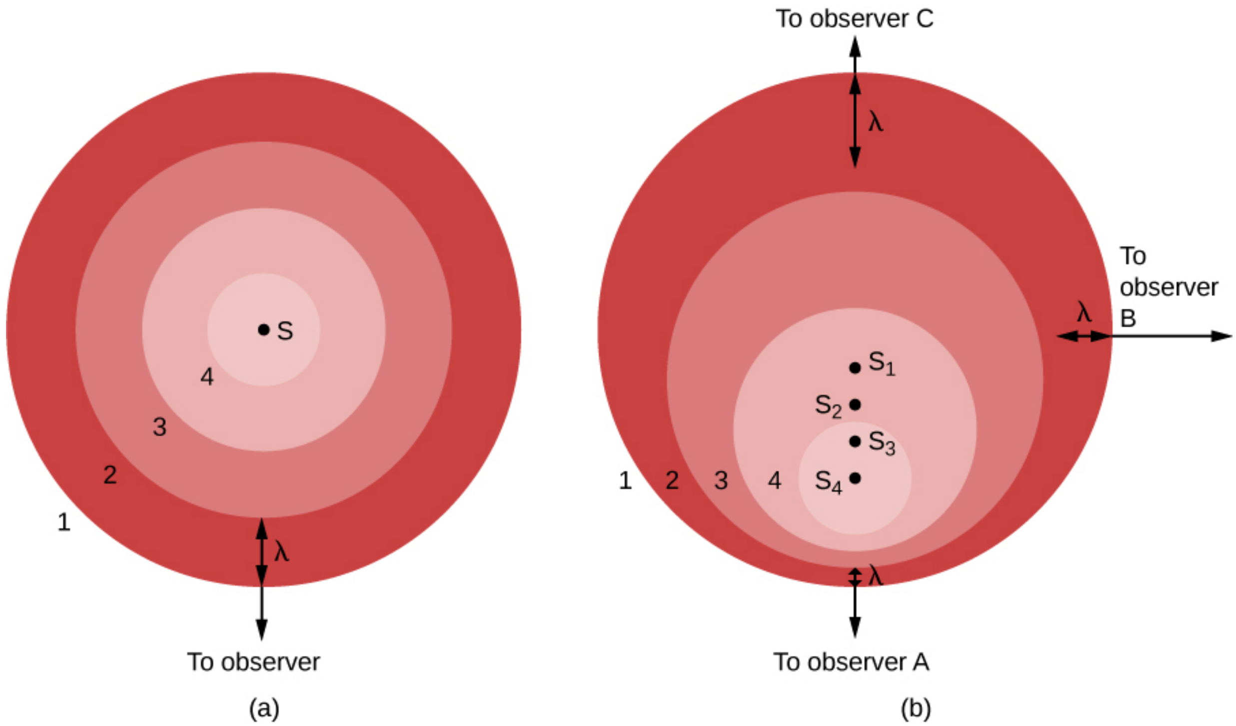
By the end of this section, you will be able to:

- Explain why the spectral lines of photons we observe from an object will change as a result of the object’s motion toward or away from us
- Describe how we can use the Doppler effect to deduce how fast astronomical objects are moving through space

The last two sections introduced you to many new concepts, and we hope that through those, you have seen one major idea emerge. Astronomers can learn about the elements in stars and galaxies by decoding the information in their spectral lines. There is a complicating factor in learning how to decode the message of starlight, however. If a star is moving toward or away from us, its lines will be in a slightly different place in the spectrum from where they would be in a star at rest. And most objects in the universe do have some motion relative to the Sun.

Motion Affects Waves

In 1842, Christian Doppler first measured the effect of motion on waves by hiring a group of musicians to play on an open railroad car as it was moving along the track. He then applied what he learned to all waves, including light, and pointed out that if a light source is approaching or receding from the observer, the light waves will be, respectively, crowded more closely together or spread out. The general principle, now known as the **Doppler effect**, is illustrated in [link].
Doppler Effect.



(a) A source, S, makes waves whose numbered crests (1, 2, 3, and 4) wash over a stationary observer. (b) The source S now moves toward observer *A* and away from observer *C*. Wave crest 1 was emitted when the source was at position S1, crest 2 at position S2, and so forth. Observer *A* sees waves compressed by this motion and sees a blueshift (if the waves are light). Observer *C* sees the waves stretched out by the motion and sees a redshift. Observer *B*, whose line of sight is perpendicular to the source's motion, sees no change in the waves (and feels left out).

In part (a) of the figure, the light source (S) is at rest with respect to the observer. The source gives off a series of waves, whose crests we have labeled 1, 2, 3, and 4. The light waves spread out evenly in all directions, like the ripples from a splash in a pond. The crests are separated by a distance, λ, where λ is the wavelength. The observer, who happens to be located in the direction of the bottom of the image, sees the light waves coming nice and evenly, one wavelength apart. Observers located anywhere else would see the same thing.

On the other hand, if the source of light is moving with respect to the observer, as seen in part (b), the situation is more complicated. Between the time one crest is emitted and the next one is ready to come out, the source has moved a bit, toward the bottom of the page. From the point of view of observer *A*, this motion of the source has decreased the distance between crests—it's squeezing the crests together, this observer might say.

In part (b), we show the situation from the perspective of three observers. The source is seen in four positions, S_1, S_2, S_3, and S_4, each corresponding to the emission of one wave crest. To observer *A*, the waves seem to follow one another more closely, at a decreased wavelength and thus increased frequency. (Remember, all light waves travel at the speed of light through empty

space, no matter what. This means that motion cannot affect the speed, but only the wavelength and the frequency. As the wavelength decreases, the frequency must increase. If the waves are shorter, more will be able to move by during each second.)

The situation is not the same for other observers. Let's look at the situation from the point of view of observer *C*, located opposite observer *A* in the figure. For her, the source is moving away from her location. As a result, the waves are not squeezed together but instead are spread out by the motion of the source. The crests arrive with an increased wavelength and decreased frequency. To observer *B*, in a direction at right angles to the motion of the source, no effect is observed. The wavelength and frequency remain the same as they were in part (a) of the figure.

We can see from this illustration that the Doppler effect is produced only by a motion toward or away from the observer, a motion called **radial velocity**. Sideways motion does not produce such an effect. Observers between *A* and *B* would observe some shortening of the light waves for that part of the motion of the source that is along their line of sight. Observers between *B* and *C* would observe lengthening of the light waves that are along their line of sight.

You may have heard the Doppler effect with sound waves. When a train whistle or police siren approaches you and then moves away, you will notice a decrease in the pitch (which is how human senses interpret sound wave frequency) of the sound waves. Compared to the waves at rest, they have changed from slightly more frequent when coming toward you, to slightly less frequent when moving away from you.

Note:
A nice example of this change in the sound of a train whistle can be heard at the end of the classic Beach Boys song "Caroline, No" on their album *Pet Sounds*. To hear this sound, go to this YouTube version of the song. The sound of the train begins at approximately 2:20.

Color Shifts

When the source of waves moves toward you, the wavelength decreases a bit. If the waves involved are visible light, then the colors of the light change slightly. As wavelength decreases, they shift toward the blue end of the spectrum: astronomers call this a *blueshift* (since the end of the spectrum is really violet, the term should probably be *violetshift*, but blue is a more common color). When the source moves away from you and the wavelength gets longer, we call the change in colors a *redshift*. Because the Doppler effect was first used with visible light in astronomy, the terms "blueshift" and "redshift" became well established. Today, astronomers use these words to describe changes in the wavelengths of radio waves or X-rays as comfortably as they use them to describe changes in visible light.

The greater the motion toward or away from us, the greater the Doppler shift. If the relative motion is entirely along the line of sight, the formula for the Doppler shift of light is
Equation:

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

where λ is the wavelength emitted by the source, Δλ is the difference between λ and the wavelength measured by the observer, *c* is the speed of light, and *v* is the relative speed of the observer and the source in the line of sight. The variable *v* is counted as positive if the velocity is one of recession, and negative if it is one of approach. Solving this equation for the velocity, we find $v = c \times \Delta\lambda/\lambda$.

If a star approaches or recedes from us, the wavelengths of light in its continuous spectrum appear shortened or lengthened, respectively, as do those of the dark lines. However, unless its speed is tens of thousands of kilometers per second, the star does not appear noticeably bluer or redder than normal. The Doppler shift is thus not easily detected in a continuous spectrum and cannot be measured accurately in such a spectrum. The wavelengths of the absorption lines can be measured accurately, however, and their Doppler shift is relatively simple to detect.

Example:

The Doppler Effect

We can use the Doppler effect equation to calculate the radial velocity of an object if we know three things: the speed of light, the original (unshifted) wavelength of the light emitted, and the difference between the wavelength of the emitted light and the wavelength we observe. For particular absorption or emission lines, we usually know exactly what wavelength the line has in our laboratories on Earth, where the source of light is not moving. We can measure the new wavelength with our instruments at the telescope, and so we know the difference in wavelength due to Doppler shifting. Since the speed of light is a universal constant, we can then calculate the radial velocity of the star.

A particular emission line of hydrogen is originally emitted with a wavelength of 656.3 nm from a gas cloud. At our telescope, we observe the wavelength of the emission line to be 656.6 nm. How fast is this gas cloud moving toward or away from Earth?

Solution

Because the light is shifted to a longer wavelength (redshifted), we know this gas cloud is moving away from us. The speed can be calculated using the Doppler shift formula:

Equation:

$$\begin{aligned} \nu &= c \times \frac{\Delta\lambda}{\lambda} = \left(3.0 \times 10^8\,\text{m/s}\right)\left(\frac{0.3\,\text{nm}}{656.3\,\text{nm}}\right) = \left(3.0 \times 10^8\,\text{m/s}\right)\left(\frac{0.3 \times 10^{-9}\,\text{m}}{656.3 \times 10^{-9}\,\text{m}}\right) \\ &= 140{,}000\,\text{m/s} = 140\,\text{km/s} \end{aligned}$$

Check Your Learning

Suppose a spectral line of hydrogen, normally at 500 nm, is observed in the spectrum of a star to be at 500.1 nm. How fast is the star moving toward or away from Earth?

Note:

Answer:

Because the light is shifted to a longer wavelength, the star is moving away from us:
$$\nu = c \times \frac{\Delta\lambda}{\lambda} = \left(3.0 \times 10^8 \text{ m/s}\right)\left(\frac{0.1 \text{ nm}}{500 \text{ nm}}\right) = \left(3.0 \times 10^8 \text{ m/s}\right)\left(\frac{0.1 \times 10^{-9} \text{ m}}{500 \times 10^{-9} \text{ m}}\right) = 60{,}000 \text{ m/s}.$$
Its speed is
60,000 m/s.

You may now be asking: if all the stars are moving and motion changes the wavelength of each spectral line, won't this be a disaster for astronomers trying to figure out what elements are present in the stars? After all, it is the precise wavelength (or color) that tells astronomers which lines belong to which element. And we first measure these wavelengths in containers of gas in our laboratories, which are not moving. If every line in a star's spectrum is now shifted by its motion to a different wavelength (color), how can we be sure which lines and which elements we are looking at in a star whose speed we do not know?

Take heart. This situation sounds worse than it really is. Astronomers rarely judge the presence of an element in an astronomical object by a single line. It is the *pattern* of lines unique to hydrogen or calcium that enables us to determine that those elements are part of the star or galaxy we are observing. The Doppler effect does not change the pattern of lines from a given element—it only shifts the whole pattern slightly toward redder or bluer wavelengths. The shifted pattern is still quite easy to recognize. Best of all, when we do recognize a familiar element's pattern, we get a bonus: the amount the pattern is shifted can enable us to determine the speed of the objects in our line of sight.

The training of astronomers includes much work on learning to decode light (and other electromagnetic radiation). A skillful "decoder" can learn the temperature of a star, what elements are in it, and even its speed in a direction toward us or away from us. That's really an impressive amount of information for stars that are light-years away.

Key Concepts and Summary

If an atom is moving toward us when an electron changes orbits and produces a spectral line, we see that line shifted slightly toward the blue of its normal wavelength in a spectrum. If the atom is moving away, we see the line shifted toward the red. This shift is known as the Doppler effect and can be used to measure the radial velocities of distant objects.

For Further Exploration

Articles

Augensen, H. & Woodbury, J. "The Electromagnetic Spectrum." *Astronomy* (June 1982): 6.

Darling, D. "Spectral Visions: The Long Wavelengths." *Astronomy* (August 1984): 16; "The Short Wavelengths." *Astronomy* (September 1984): 14.

Gingerich, O. “Unlocking the Chemical Secrets of the Cosmos.” *Sky & Telescope* (July 1981): 13.

Stencil, R. et al. “Astronomical Spectroscopy.” *Astronomy* (June 1978): 6.

Websites

Doppler Effect: http://www.physicsclassroom.com/class/waves/Lesson-3/The-Doppler-Effect. A shaking bug and the Doppler Effect explained.

Electromagnetic Spectrum: http://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html. An introduction to the electromagnetic spectrum from NASA’s *Imagine the Universe*; note that you can click the “Advanced” button near the top and get a more detailed discussion.

Rainbows: How They Form and How to See Them: http://www.livescience.com/30235-rainbows-formation-explainer.html. By meteorologist and amateur astronomer Joe Rao.

Videos

Doppler Effect: http://www.esa.int/spaceinvideos/Videos/2014/07/Doppler_effect_-_classroom_demonstration_video_VP05. ESA video with Doppler ball demonstration and Doppler effect and satellites (4:48).

How a Prism Works to Make Rainbow Colors: https://www.youtube.com/watch?v=JGqsi_LDUn0. Short video on how a prism bends light to make a rainbow of colors (2:44).

Tour of the Electromagnetic Spectrum: https://www.youtube.com/watch?v=HPcAWNlVl-8. *NASA Mission Science* video tour of the bands of the electromagnetic spectrum (eight short videos).

Introductions to Quantum Mechanics

Ford, Kenneth. *The Quantum World.* 2004. A well-written recent introduction by a physicist/educator.

Gribbin, John. *In Search of Schroedinger’s Cat.* 1984. Clear, very basic introduction to the fundamental ideas of quantum mechanics, by a British physicist and science writer.

Rae, Alastair. *Quantum Physics: A Beginner’s Guide.* 2005. Widely praised introduction by a British physicist.

Collaborative Group Activities

A. Have your group make a list of all the electromagnetic wave technology you use during a typical day.

B. How many applications of the Doppler effect can your group think of in everyday life? For example, why would the highway patrol find it useful?
C. Have members of your group go home and "read" the face of your radio set and then compare notes. If you do not have a radio, research "broadcast radio frequencies" to find answers to the following questions. What do all the words and symbols mean? What frequencies can your radio tune to? What is the frequency of your favorite radio station? What is its wavelength?
D. If your instructor were to give you a spectrometer, what kind of spectra does your group think you would see from each of the following: (1) a household lightbulb, (2) the Sun, (3) the "neon lights of Broadway," (4) an ordinary household flashlight, and (5) a streetlight on a busy shopping street?
E. Suppose astronomers want to send a message to an alien civilization that is living on a planet with an atmosphere very similar to that of Earth's. This message must travel through space, make it through the other planet's atmosphere, and be noticeable to the residents of that planet. Have your group discuss what band of the electromagnetic spectrum might be best for this message and why. (Some people, including noted physicist Stephen Hawking, have warned scientists not to send such messages and reveal the presence of our civilization to a possible hostile cosmos. Do you agree with this concern?)

Review Questions

Exercise:

Problem:

What distinguishes one type of electromagnetic radiation from another? What are the main categories (or bands) of the electromagnetic spectrum?

Exercise:

Problem: What is a wave? Use the terms *wavelength* and *frequency* in your definition.

Exercise:

Problem:

Is your textbook the kind of idealized object (described in section on radiation laws) that absorbs all the radiation falling on it? Explain. How about the black sweater worn by one of your classmates?

Exercise:

Problem: Where in an atom would you expect to find electrons? Protons? Neutrons?

Exercise:

Problem:

Explain how emission lines and absorption lines are formed. In what sorts of cosmic objects would you expect to see each?

Exercise:

Problem:

Explain how the Doppler effect works for sound waves and give some familiar examples.

Exercise:

Problem: What kind of motion for a star does not produce a Doppler effect? Explain.

Exercise:

Problem: Describe how Bohr's model used the work of Maxwell.

Exercise:

Problem: Explain why light is referred to as electromagnetic radiation.

Exercise:

Problem:

Explain the difference between radiation as it is used in most everyday language and radiation as it is used in an astronomical context.

Exercise:

Problem: What are the differences between light waves and sound waves?

Exercise:

Problem:

Which type of wave has a longer wavelength: AM radio waves (with frequencies in the kilohertz range) or FM radio waves (with frequencies in the megahertz range)? Explain.

Exercise:

Problem:

Explain why astronomers long ago believed that space must be filled with some kind of substance (the "aether") instead of the vacuum we know it is today.

Exercise:

Problem: Explain what the ionosphere is and how it interacts with some radio waves.

Exercise:

Problem: Which is more dangerous to living things, gamma rays or X-rays? Explain.

Exercise:

Problem:

Explain why we have to observe stars and other astronomical objects from above Earth's atmosphere in order to fully learn about their properties.

Exercise:

Problem:

Explain why hotter objects tend to radiate more energetic photons compared to cooler objects.

Exercise:

Problem: Explain how we can deduce the temperature of a star by determining its color.

Exercise:

Problem:

Explain what dispersion is and how astronomers use this phenomenon to study a star's light.

Exercise:

Problem: Explain why glass prisms disperse light.

Exercise:

Problem: Explain what Joseph Fraunhofer discovered about stellar spectra.

Exercise:

Problem:

Explain how we use spectral absorption and emission lines to determine the composition of a gas.

Exercise:

Problem:

Explain the results of Rutherford's gold foil experiment and how they changed our model of the atom.

Exercise:

Problem:

Is it possible for two different atoms of carbon to have different numbers of neutrons in their nuclei? Explain.

Exercise:

Problem: What are the three isotopes of hydrogen, and how do they differ?

Exercise:

Problem:

Explain how electrons use light energy to move among energy levels within an atom.

Exercise:

Problem:

Explain why astronomers use the term “blueshifted” for objects moving toward us and “redshifted” for objects moving away from us.

Exercise:

Problem:

If spectral line wavelengths are changing for objects based on the radial velocities of those objects, how can we deduce which type of atom is responsible for a particular absorption or emission line?

Thought Questions

Exercise:

Problem:

Make a list of some of the many practical consequences of Maxwell’s theory of electromagnetic waves (television is one example).

Exercise:

Problem: With what type of electromagnetic radiation would you observe:

A. A star with a temperature of 5800 K?
B. A gas heated to a temperature of one million K?
C. A person on a dark night?

Exercise:

Problem:

Why is it dangerous to be exposed to X-rays but not (or at least much less) dangerous to be exposed to radio waves?

Exercise:

Problem:

Go outside on a clear night, wait 15 minutes for your eyes to adjust to the dark, and look carefully at the brightest stars. Some should look slightly red and others slightly blue. The primary factor that determines the color of a star is its temperature. Which is hotter: a blue star or a red one? Explain

Exercise:

Problem:

Water faucets are often labeled with a red dot for hot water and a blue dot for cold. Given Wien's law, does this labeling make sense?

Exercise:

Problem:

Suppose you are standing at the exact center of a park surrounded by a circular road. An ambulance drives completely around this road, with siren blaring. How does the pitch of the siren change as it circles around you?

Exercise:

Problem:

How could you measure Earth's orbital speed by photographing the spectrum of a star at various times throughout the year? (Hint: Suppose the star lies in the plane of Earth's orbit.)

Exercise:

Problem:

Astronomers want to make maps of the sky showing sources of X-rays or gamma rays. Explain why those X-rays and gamma rays must be observed from above Earth's atmosphere.

Exercise:

Problem:

The greenhouse effect can be explained easily if you understand the laws of blackbody radiation. A greenhouse gas blocks the transmission of infrared light. Given that the incoming light to Earth is sunlight with a characteristic temperature of 5800 K (which peaks in the visible part of the spectrum) and the outgoing light from Earth has a characteristic temperature of about 300 K (which peaks in the infrared part of the spectrum), explain how greenhouse gases cause Earth to warm up. As part of your answer, discuss that greenhouse gases block both incoming and outgoing infrared light. Explain why these two effects don't simply cancel each other, leading to no net temperature change.

Exercise:

Problem:

An idealized radiating object does not reflect or scatter any radiation but instead absorbs all of the electromagnetic energy that falls on it. Can you explain why astronomers call such an object a blackbody? Keep in mind that even stars, which shine brightly in a variety of colors, are considered blackbodies. Explain why.

Exercise:

Problem:

Why are ionized gases typically only found in very high-temperature environments?

Exercise:

Problem: Explain why each element has a unique spectrum of absorption or emission lines.

Figuring for Yourself

Exercise:

Problem:

What is the wavelength of the carrier wave of a campus radio station, broadcasting at a frequency of 97.2 MHz (million cycles per second or million hertz)?

Exercise:

Problem:

What is the frequency of a red laser beam, with a wavelength of 670 nm, which your astronomy instructor might use to point to slides during a lecture on galaxies?

Exercise:

Problem:

You go to a dance club to forget how hard your astronomy midterm was. What is the frequency of a wave of ultraviolet light coming from a blacklight in the club, if its wavelength is 150 nm?

Exercise:

Problem: What is the energy of the photon with the frequency you calculated in [link]?

Exercise:

Problem:

If the emitted infrared radiation from Pluto, has a wavelength of maximum intensity at 75,000 nm, what is the temperature of Pluto assuming it follows Wien’s law?

Exercise:

Problem:

What is the temperature of a star whose maximum light is emitted at a wavelength of 290 nm?

Glossary

Doppler effect

the apparent change in wavelength or frequency of the radiation from a source due to its relative motion away from or toward the observer

radial velocity

motion toward or away from the observer; the component of relative velocity that lies in the line of sight

Thinking Ahead
class="introduction"
Hubble Space Telescope (HST).

This artist's impression shows the Hubble above Earth, with the rectangular solar panels that provide it with power seen to the left and right.

If you look at the sky when you are far away from city lights, there seem to be an overwhelming number of stars up there. In reality, only about 9000

stars are visible to the unaided eye (from both hemispheres of our planet). The light from most stars is so weak that by the time it reaches Earth, it cannot be detected by the human eye. How can we learn about the vast majority of objects in the universe that our unaided eyes simply cannot see?

In this chapter, we describe the tools astronomers use to extend their vision into space. We have learned almost everything we know about the universe from studying electromagnetic radiation, as discussed in the chapter on Radiation and Spectra. In the twentieth century, our exploration of space made it possible to detect electromagnetic radiation at all wavelengths, from gamma rays to radio waves. The different wavelengths carry different kinds of information, and the appearance of any given object often depends on the wavelength at which the observations are made.

Telescopes

LEARNING OBJECTIVES

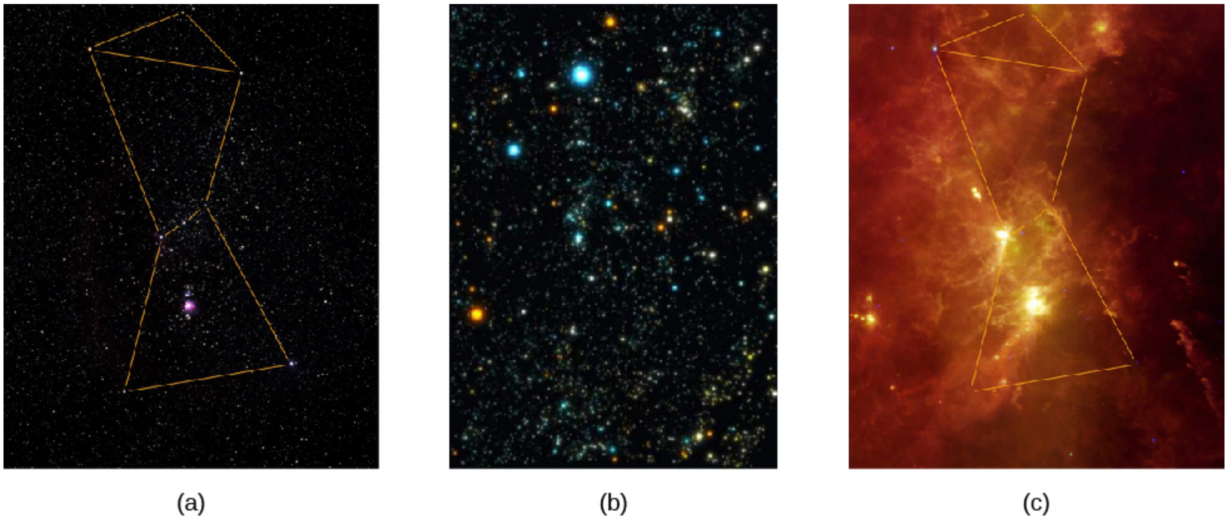
By the end of this section, you will be able to:

- Describe the three basic components of a modern system for measuring astronomical sources
- Describe the main functions of a telescope
- Describe the two basic types of visible-light telescopes and how they form images

Systems for Measuring Radiation

There are three basic components of a modern system for measuring radiation from astronomical sources. First, there is a **telescope**, which serves as a “bucket” for collecting visible light (or radiation at other wavelengths, as shown in ([link]). Just as you can catch more rain with a garbage can than with a coffee cup, large telescopes gather much more light than your eye can. Second, there is an instrument attached to the telescope that sorts the incoming radiation by wavelength. Sometimes the sorting is fairly crude. For example, we might simply want to separate blue light from red light so that we can determine the temperature of a star. But at other times, we want to see individual spectral lines to determine what an object is made of, or to measure its speed (as explained in the Radiation and Spectra chapter). Third, we need some type of **detector**, a device that senses the radiation in the wavelength regions we have chosen and permanently records the observations.

Orion Region at Different Wavelengths.



The same part of the sky looks different when observed with instruments that are sensitive to different bands of the spectrum. (a) Visible light: this shows part of the Orion region as the human eye sees it, with dotted lines added to show the figure of the mythical hunter, Orion. (b) X-rays: here, the view emphasizes the point-like X-ray sources nearby. The colors are artificial, changing from yellow to white to blue with increasing energy of the X-rays. The bright, hot stars in Orion are still seen in this image, but so are many other objects located at very different distances, including other stars, star corpses, and galaxies at the edge of the observable universe. (c) Infrared radiation: here, we mainly see the glowing dust in this region. (credit a: modification of work by Howard McCallon/NASA/IRAS; credit b: modification of work by Howard McCallon/NASA/IRAS; credit c: modification of work by Michael F. Corcoran)

The history of the development of astronomical telescopes is about how new technologies have been applied to improve the efficiency of these three basic components: the telescopes, the wavelength-sorting device, and the detectors. Let’s first look at the development of the telescope.

Many ancient cultures built special sites for observing the sky ([link]). At these ancient *observatories*, they could measure the positions of celestial objects, mostly to keep track of time and date. Many of these ancient

observatories had religious and ritual functions as well. The eye was the only device available to gather light, all of the colors in the light were observed at once, and the only permanent record of the observations was made by human beings writing down or sketching what they saw.

Two Pre-Telescopic Observatories.

(a) (b)

(a) Machu Picchu is a fifteenth century Incan site located in Peru. (b) Stonehenge, a prehistoric site (3000–2000 BCE), is located in England. (credit a: modification of work by Allard Schmidt)

While Hans Lippershey, Zaccharias Janssen, and Jacob Metius are all credited with the invention of the telescope around 1608—applying for patents within weeks of each other—it was Galileo who, in 1610, used this simple tube with lenses (which he called a spyglass) to observe the sky and gather more light than his eyes alone could. Even his small telescope—used over many nights—revolutionized ideas about the nature of the planets and the position of Earth.

How Telescopes Work

Telescopes have come a long way since Galileo’s time. Now they tend to be huge devices; the most expensive cost hundreds of millions to billions of dollars. (To provide some reference point, however, keep in mind that just renovating college football stadiums typically costs hundreds of millions of

dollars—with the most expensive recent renovation, at Texas A&M University's Kyle Field, costing $450 million.) The reason astronomers keep building bigger and bigger telescopes is that celestial objects—such as planets, stars, and galaxies—send much more light to Earth than any human eye (with its tiny opening) can catch, and bigger telescopes can detect fainter objects. If you have ever watched the stars with a group of friends, you know that there's plenty of starlight to go around; each of you can see each of the stars. If a thousand more people were watching, each of them would also catch a bit of each star's light. Yet, as far as you are concerned, the light not shining into your eye is wasted. It would be great if some of this "wasted" light could also be captured and brought to your eye. This is precisely what a telescope does.

The most important functions of a telescope are (1) to *collect* the faint light from an astronomical source and (2) to *focus* all the light into a point or an image. Most objects of interest to astronomers are extremely faint: the more light we can collect, the better we can study such objects. (And remember, even though we are focusing on visible light first, there are many telescopes that collect other kinds of electromagnetic radiation.)

Telescopes that collect visible radiation use a lens or mirror to gather the light. Other types of telescopes may use collecting devices that look very different from the lenses and mirrors with which we are familiar, but they serve the same function. In all types of telescopes, the light-gathering ability is determined by the area of the device acting as the light-gathering "bucket." Since most telescopes have mirrors or lenses, we can compare their light-gathering power by comparing the **apertures**, or diameters, of the opening through which light travels or reflects.

The amount of light a telescope can collect increases with the size of the aperture. A telescope with a mirror that is 4 meters in diameter can collect 16 times as much light as a telescope that is 1 meter in diameter. (The diameter is squared because the area of a circle equals $\pi d^2/4$, where d is the diameter of the circle.)

Example:

Calculating the Light-Collecting Area

What is the area of a 1-m diameter telescope? A 4-m diameter one?

Solution

Using the equation for the area of a circle,

Equation:

$$A = \frac{\pi d^2}{4}$$

the area of a 1-m telescope is

Equation:

$$\frac{\pi d^2}{4} = \frac{\pi (1\ \mathrm{m})^2}{4} = 0.79\ \mathrm{m}^2$$

and the area of a 4-m telescope is

Equation:

$$\frac{\pi d^2}{4} = \frac{\pi (4\ \mathrm{m})^2}{4} = 12.6\ \mathrm{m}^2$$

Check Your Learning

Show that the ratio of the two areas is 16:1.

Note:

Answer:

$\frac{12.6\ \mathrm{m}^2}{0.79\ \mathrm{m}^2} = 16.$ Therefore, with 16 times the area, a 4-m telescope collects 16 times the light of a 1-m telescope.

After the telescope forms an image, we need some way to detect and record it so that we can measure, reproduce, and analyze the image in various

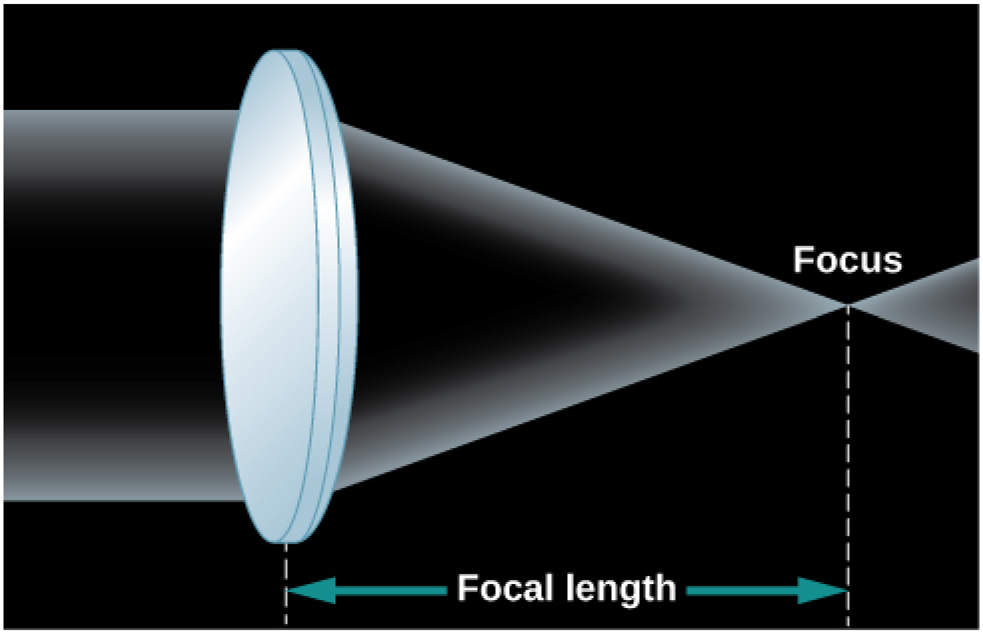
ways. Before the nineteenth century, astronomers simply viewed images with their eyes and wrote descriptions of what they saw. This was very inefficient and did not lead to a very reliable long-term record; you know from crime shows on television that eyewitness accounts are often inaccurate.

In the nineteenth century, the use of photography became widespread. In those days, photographs were a chemical record of an image on a specially treated glass plate. Today, the image is generally detected with sensors similar to those in digital cameras, recorded electronically, and stored in computers. This permanent record can then be used for detailed and quantitative studies. Professional astronomers rarely look through the large telescopes that they use for their research.

Formation of an Image by a Lens or a Mirror

Whether or not you wear glasses, you see the world through lenses; they are key elements of your eyes. A lens is a transparent piece of material that bends the rays of light passing through it. If the light rays are parallel as they enter, the lens brings them together in one place to form an image ([link]). If the curvatures of the lens surfaces are just right, all parallel rays of light (say, from a star) are bent, or *refracted*, in such a way that they converge toward a point, called the **focus** of the lens. At the focus, an image of the light source appears. In the case of parallel light rays, the distance from the lens to the location where the light rays focus, or image, behind the lens is called the *focal length* of the lens.

Formation of an Image by a Simple Lens.



Parallel rays from a distant source are bent by the convex lens so that they all come together in a single place (the focus) to form an image.

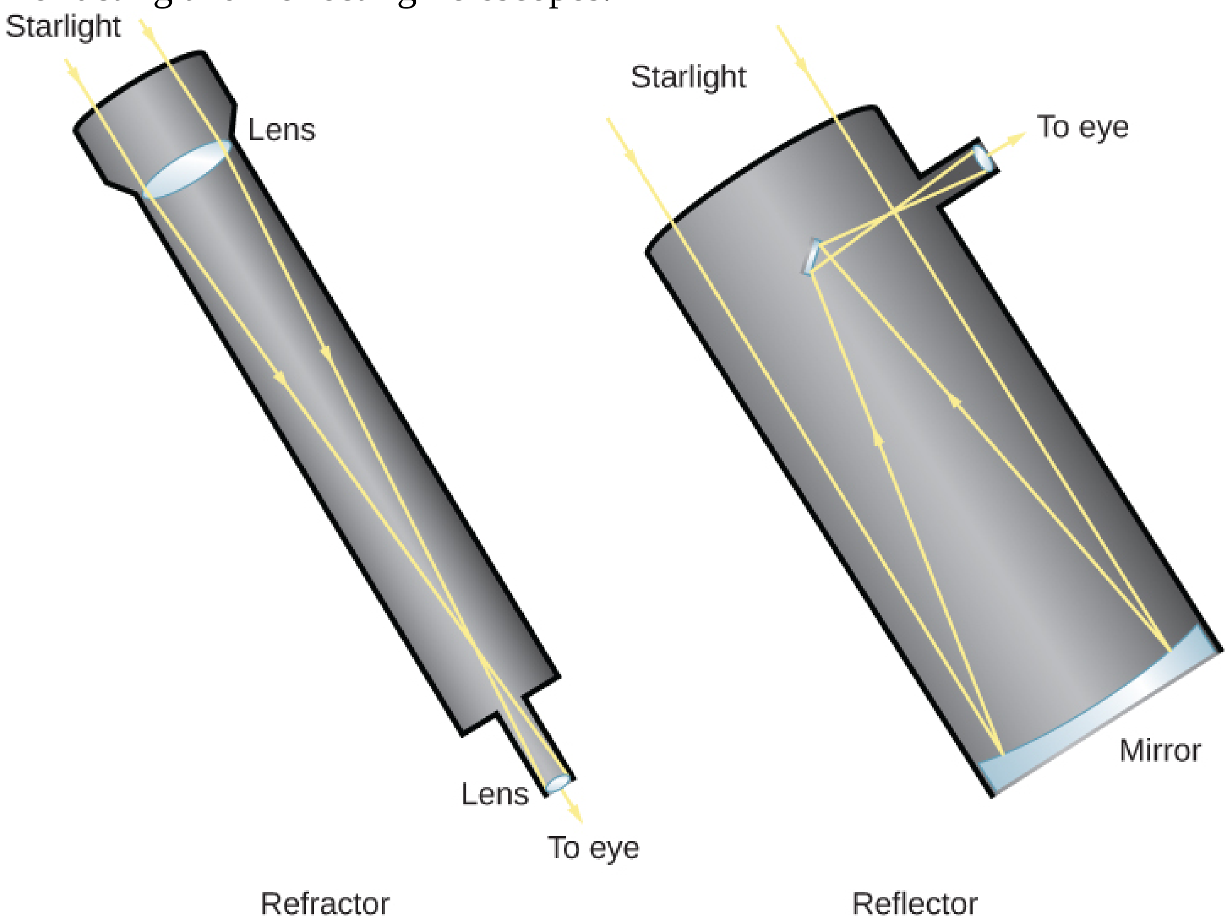
As you look at [link], you may ask why two rays of light from the same star would be parallel to each other. After all, if you draw a picture of star shining in all directions, the rays of light coming from the star don’t look parallel at all. But remember that the stars (and other astronomical objects) are all extremely far away. By the time the few rays of light pointed toward us actually arrive at Earth, they are, for all practical purposes, parallel to each other. Put another way, any rays that were *not* parallel to the ones pointed at Earth are now heading in some very different direction in the universe.

To view the image formed by the lens in a telescope, we use an additional lens called an **eyepiece**. The eyepiece focuses the image at a distance that is either directly viewable by a human or at a convenient place for a detector. Using different eyepieces, we can change the *magnification* (or size) of the image and also redirect the light to a more accessible location. Stars look like points of light, and magnifying them makes little difference, but the

image of a planet or a galaxy, which has structure, can often benefit from being magnified.

Many people, when thinking of a telescope, picture a long tube with a large glass lens at one end. This design, which uses a lens as its main optical element to form an image, as we have been discussing, is known as a *refractor* ([link]), and a telescope based on this design is called a **refracting telescope**. Galileo's telescopes were refractors, as are today's binoculars and field glasses. However, there is a limit to the size of a refracting telescope. The largest one ever built was a 49-inch refractor built for the Paris 1900 Exposition, and it was dismantled after the Exposition. Currently, the largest refracting telescope is the 40-inch refractor at Yerkes Observatory in Wisconsin.

Refracting and Reflecting Telescopes.



Light enters a refracting telescope through a lens at the upper end, which focuses the light near the bottom of the telescope. An eyepiece

then magnifies the image so that it can be viewed by the eye, or a detector like a photographic plate can be placed at the focus. The upper end of a reflecting telescope is open, and the light passes through to the mirror located at the bottom of the telescope. The mirror then focuses the light at the top end, where it can be detected. Alternatively, as in this sketch, a second mirror may reflect the light to a position outside the telescope structure, where an observer can have easier access to it. Professional astronomers' telescopes are more complicated than this, but they follow the same principles of reflection and refraction.

One problem with a refracting telescope is that the light must pass *through* the lens of a refractor. That means the glass must be perfect all the way through, and it has proven very difficult to make large pieces of glass without flaws and bubbles in them. Also, optical properties of transparent materials change a little bit with the wavelengths (or colors) of light, so there is some additional distortion, known as **chromatic aberration**. Each wavelength focuses at a slightly different spot, causing the image to appear blurry.

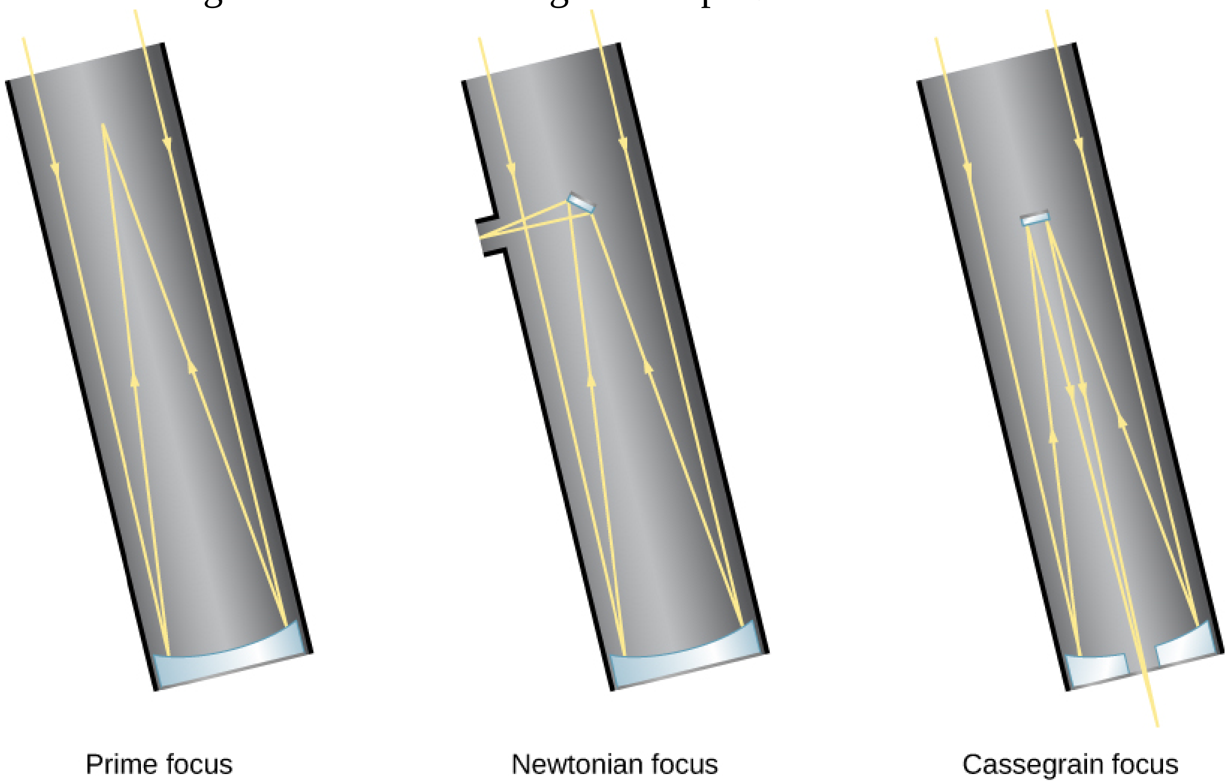
In addition, since the light must pass through the lens, the lens can only be supported around its edges (just like the frames of our eyeglasses). The force of gravity will cause a large lens to sag and distort the path of the light rays as they pass through it. Finally, because the light passes through it, both sides of the lens must be manufactured to precisely the right shape in order to produce a sharp image.

A different type of telescope uses a concave *primary mirror* as its main optical element. The mirror is curved like the inner surface of a sphere, and it reflects light in order to form an image ([link]). Telescope mirrors are coated with a shiny metal, usually silver, aluminum, or, occasionally, gold, to make them highly reflective. If the mirror has the correct shape, all parallel rays are reflected back to the same point, the focus of the mirror. Thus, images are produced by a mirror exactly as they are by a lens.

Telescopes designed with mirrors avoid the problems of refracting telescopes. Because the light is reflected from the front surface only, flaws and bubbles within the glass do not affect the path of the light. In a telescope designed with mirrors, only the front surface has to be manufactured to a precise shape, and the mirror can be supported from the back. For these reasons, most astronomical telescopes today (both amateur and professional) use a mirror rather than a lens to form an image; this type of telescope is called a **reflecting telescope**. The first successful reflecting telescope was built by Isaac Newton in 1668.

In a reflecting telescope, the concave mirror is placed at the bottom of a tube or open framework. The mirror reflects the light back up the tube to form an image near the front end at a location called the **prime focus**. The image can be observed at the prime focus, or additional mirrors can intercept the light and redirect it to a position where the observer can view it more easily ([link]). Since an astronomer at the prime focus can block much of the light coming to the main mirror, the use of a small *secondary mirror* allows more light to get through the system.

Focus Arrangements for Reflecting Telescopes.



Reflecting telescopes have different options for where the light is brought to a focus. With prime focus, light is detected where it comes to a focus after reflecting from the primary mirror. With Newtonian focus, light is reflected by a small secondary mirror off to one side, where it can be detected (see also [link]). Most large professional telescopes have a Cassegrain focus in which light is reflected by the secondary mirror down through a hole in the primary mirror to an observing station below the telescope.

Note:

Choosing Your Own Telescope

If the astronomy course you are taking whets your appetite for exploring the sky further, you may be thinking about buying your own telescope. Many excellent amateur telescopes are available, and some research is required to find the best model for your needs. Some good sources of information about personal telescopes are the two popular US magazines aimed at amateur astronomers: *Sky & Telescope* and *Astronomy*. Both carry regular articles with advice, reviews, and advertisements from reputable telescope dealers.

Some of the factors that determine which telescope is right for you depend upon your preferences:

- Will you be setting up the telescope in one place and leaving it there, or do you want an instrument that is portable and can come with you on outdoor excursions? How portable should it be, in terms of size and weight?
- Do you want to observe the sky with your eyes only, or do you want to take photographs? (Long-exposure photography, for example, requires a good clock drive to turn your telescope to compensate for Earth’s rotation.)
- What types of objects will you be observing? Are you interested primarily in comets, planets, star clusters, or galaxies, or do you want to observe all kinds of celestial sights?

You may not know the answers to some of these questions yet. For this reason, you may want to “test-drive” some telescopes first. Most communities have amateur astronomy clubs that sponsor star parties open to the public. The members of those clubs often know a lot about telescopes and can share their ideas with you. Your instructor may know where the nearest amateur astronomy club meets; or, to find a club near you, use the websites suggested in Appendix B.

Furthermore, you may already have an instrument like a telescope at home (or have access to one through a relative or friend). Many amateur astronomers recommend starting your survey of the sky with a good pair of binoculars. These are easily carried around and can show you many objects not visible (or clear) to the unaided eye.

When you are ready to purchase a telescope, you might find the following ideas useful:

- The key characteristic of a telescope is the aperture of the main mirror or lens; when someone says they have a 6-inch or 8-inch telescope, they mean the diameter of the collecting surface. The larger the aperture, the more light you can gather, and the fainter the objects you can see or photograph.
- Telescopes of a given aperture that use lenses (refractors) are typically more expensive than those using mirrors (reflectors) because both sides of a lens must be polished to great accuracy. And, because the light passes through it, the lens must be made of high-quality glass throughout. In contrast, only the front surface of a mirror must be accurately polished.
- Magnification is not one of the criteria on which to base your choice of a telescope. As we discussed, the magnification of the image is done by a smaller eyepiece, so the magnification can be adjusted by changing eyepieces. However, a telescope will magnify not only the astronomical object you are viewing but also the turbulence of Earth’s atmosphere. If the magnification is too high, your image will shimmer and shake and be difficult to view. A good telescope will come with a variety of eyepieces that stay within the range of useful magnification.
- The mount of a telescope (the structure on which it rests) is one of its most critical elements. Because a telescope shows a tiny field of view, which is magnified significantly, even the smallest vibration or jarring

of the telescope can move the object you are viewing around or out of your field of view. A sturdy and stable mount is essential for serious viewing or photography (although it clearly affects how portable your telescope can be).
- A telescope requires some practice to set up and use effectively. Don't expect everything to go perfectly on your first try. Take some time to read the instructions. If a local amateur astronomy club is nearby, use it as a resource.

A telescope collects the faint light from astronomical sources and brings it to a focus, where an instrument can sort the light according to wavelength. Light is then directed to a detector, where a permanent record is made. The light-gathering power of a telescope is determined by the diameter of its aperture, or opening—that is, by the area of its largest or primary lens or mirror. The primary optical element in a telescope is either a convex lens (in a refracting telescope) or a concave mirror (in a reflector) that brings the light to a focus. Most large telescopes are reflectors; it is easier to manufacture and support large mirrors because the light does not have to pass through glass.

Glossary

aperture
: diameter of the primary lens or mirror of a telescope

chromatic aberration
: distortion that causes an image to appear fuzzy when each wavelength coming into a transparent material focuses at a different spot

detector
: device sensitive to electromagnetic radiation that makes a record of astronomical observations

eyepiece

magnifying lens used to view the image produced by the objective lens or primary mirror of a telescope

focus

(of telescope) point where the rays of light converged by a mirror or lens meet

prime focus

point in a telescope where the objective lens or primary mirror focuses the light

reflecting telescope

telescope in which the principal light collector is a concave mirror

refracting telescope

telescope in which the principal light collector is a lens or system of lenses

telescope

instrument for collecting visible-light or other electromagnetic radiation

Visible-Light Detectors and Instruments

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the difference between photographic plates and charge-coupled devices
- Describe the unique difficulties associated with infrared observations and their solutions
- Describe how a spectrometer works

After a telescope collects radiation from an astronomical source, the radiation must be *detected* and measured. The first detector used for astronomical observations was the human eye, but it suffers from being connected to an imperfect recording and retrieving device—the human brain. Photography and modern electronic detectors have eliminated the quirks of human memory by making a permanent record of the information from the cosmos.

The eye also suffers from having a very short *integration time*; it takes only a fraction of a second to add light energy together before sending the image to the brain. One important advantage of modern detectors is that the light from astronomical objects can be collected by the detector over longer periods of time; this technique is called “taking a long exposure.” Exposures of several hours are required to detect very faint objects in the cosmos.

Before the light reaches the detector, astronomers today normally use some type of instrument to sort the light according to wavelength. The instrument may be as simple as colored filters, which transmit light within a specified range of wavelengths. A red transparent plastic is an everyday example of a filter that transmits only the red light and blocks the other colors. After the light passes through a filter, it forms an image that astronomers can then use to measure the apparent brightness and color of objects. We will show you many examples of such images in the later chapters of this book, and we will describe what we can learn from them.

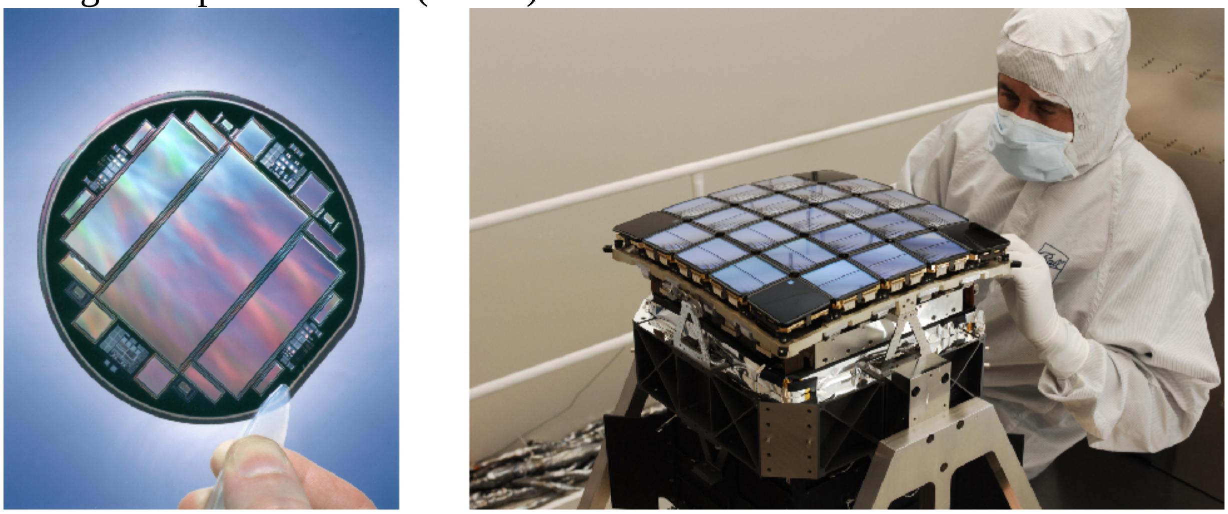
Alternatively, the instrument between telescope and detector may be one of several devices that spread the light out into its full rainbow of colors so that astronomers can measure individual lines in the spectrum. Such an instrument (which you learned about in the chapter on Radiation and Spectra) is called a *spectrometer* because it allows astronomers to measure (to meter) the spectrum of a source of radiation. Whether a filter or a spectrometer, both types of wavelength-sorting instruments still have to use detectors to record and measure the properties of light.

Photographic and Electronic Detectors

Throughout most of the twentieth century, photographic film or *glass plates* served as the prime astronomical detectors, whether for photographing spectra or direct images of celestial objects. In a photographic plate, a light-sensitive chemical coating is applied to a piece of glass that, when developed, provides a lasting record of the image. At observatories around the world, vast collections of photographs preserve what the sky has looked like during the past 100 years. Photography represents a huge improvement over the human eye, but it still has limitations. Photographic films are inefficient: only about 1% of the light that actually falls on the film contributes to the chemical change that makes the image; the rest is wasted.

Astronomers today have much more efficient electronic detectors to record astronomical images. Most often, these are **charge-coupled devices (CCDs)**, which are similar to the detectors used in video camcorders or in digital cameras (like the one more and more students have on their cell phones) (see [link]). In a CCD, photons of radiation hitting any part of the

detector generate a stream of charged particles (electrons) that are stored and counted at the end of the exposure. Each place where the radiation is counted is called a pixel (picture element), and modern detectors can count the photons in millions of pixels (megapixels, or MPs).
Charge-Coupled Devices (CCDs).



(a) (b)

(a) This CCD is a mere 300-micrometers thick (thinner than a human hair) yet holds more than 21 million pixels. (b) This matrix of 42 CCDs serves the Kepler telescope. (credit a: modification of work by US Department of Energy; credit b: modification of work by NASA and Ball Aerospace)

Because CCDs typically record as much as 60–70% of all the photons that strike them, and the best silicon and infrared CCDs exceed 90% sensitivity, we can detect much fainter objects. Among these are many small moons around the outer planets, icy dwarf planets beyond Pluto, and dwarf galaxies of stars. CCDs also provide more accurate measurements of the brightness of astronomical objects than photography, and their output is digital—in the form of numbers that can go directly into a computer for analysis.

Infrared Observations

Observing the universe in the infrared band of the spectrum presents some additional challenges. The infrared region extends from wavelengths near 1 micrometer (µm), which is about the long wavelength sensitivity limit of both CCDs and photography, to 100 micrometers or longer. Recall from the discussion on radiation and spectra that infrared is “heat radiation” (given off at temperatures that we humans are comfortable with). The main challenge to astronomers using infrared is to distinguish between the tiny amount of heat radiation that reaches Earth from stars and galaxies, and the much greater heat radiated by the telescope itself and our planet’s atmosphere.

Typical temperatures on Earth’s surface are near 300 K, and the atmosphere through which observations are made is only a little cooler. According to Wien’s law (from the chapter on Radiation and Spectra), the telescope, the observatory, and even the sky are radiating infrared energy with a peak wavelength of about 10 micrometers. To infrared eyes, everything on Earth is brightly aglow—including the telescope and camera ([link]). The challenge is to detect faint cosmic sources against this sea of infrared light. Another way to look at this is that an astronomer using infrared must always contend with the situation that a visible-light observer would face if working in broad daylight with a telescope and optics lined with bright fluorescent lights.
Infrared Eyes.

Infrared waves can penetrate places in the universe from which light is blocked, as shown in this infrared image where the plastic bag blocks visible light but not infrared. (credit: NASA/JPL-Caltech/R. Hurt (SSC))

To solve this problem, astronomers must protect the infrared detector from nearby radiation, just as you would shield photographic film from bright daylight. Since anything warm radiates infrared energy, the detector must be isolated in very cold surroundings; often, it is held near absolute zero (1 to 3 K) by immersing it in liquid helium. The second step is to reduce the radiation emitted by the telescope structure and optics, and to block this heat from reaching the infrared detector.

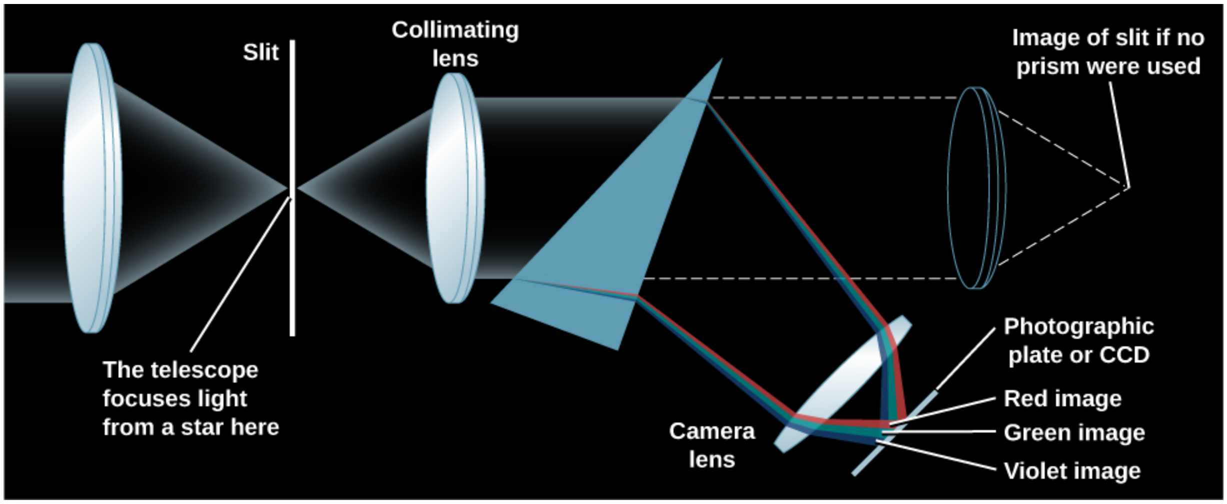
Note:
Check out The Infrared Zoo to get a sense of what familiar objects look like with infrared radiation. Slide the slider to change the wavelength of

radiation for the picture, and click the arrow to see other animals.

Spectroscopy

Spectroscopy is one of the astronomer’s most powerful tools, providing information about the composition, temperature, motion, and other characteristics of celestial objects. More than half of the time spent on most large telescopes is used for spectroscopy.

The many different wavelengths present in light can be separated by passing them through a spectrometer to form a spectrum. The design of a simple spectrometer is illustrated in [link]. Light from the source (actually, the image of a source produced by the telescope) enters the instrument through a small hole or narrow slit, and is collimated (made into a beam of parallel rays) by a lens. The light then passes through a prism, producing a spectrum: different wavelengths leave the prism in different directions because each wavelength is bent by a different amount when it enters and leaves the prism. A second lens placed behind the prism focuses the many different images of the slit or entrance hole onto a CCD or other detecting device. This collection of images (spread out by color) is the spectrum that astronomers can then analyze at a later point. As spectroscopy spreads the light out into more and more collecting bins, fewer photons go into each bin, so either a larger telescope is needed or the integration time must be greatly increased—usually both.
Prism Spectrometer.



The light from the telescope is focused on a slit. A prism (or grating) disperses the light into a spectrum, which is then photographed or recorded electronically.

In practice, astronomers today are more likely to use a different device, called a *grating*, to disperse the spectrum. A grating is a piece of material with thousands of grooves on its surface. While it functions completely differently, a grating, like a prism, also spreads light out into a spectrum.

Visible-light detectors include the human eye, photographic film, and charge-coupled devices (CCDs). Detectors that are sensitive to infrared radiation must be cooled to very low temperatures since everything in and near the telescope gives off infrared waves. A spectrometer disperses the light into a spectrum to be recorded for detailed analysis.

Glossary

charge-coupled device (CCD)
: array of high-sensitivity electronic detectors of electromagnetic radiation, used at the focus of a telescope (or camera lens) to record an image or spectrum

Observations outside Earth’s Atmosphere

LEARNING OBJECTIVES

By the end of this section, you will be able to:

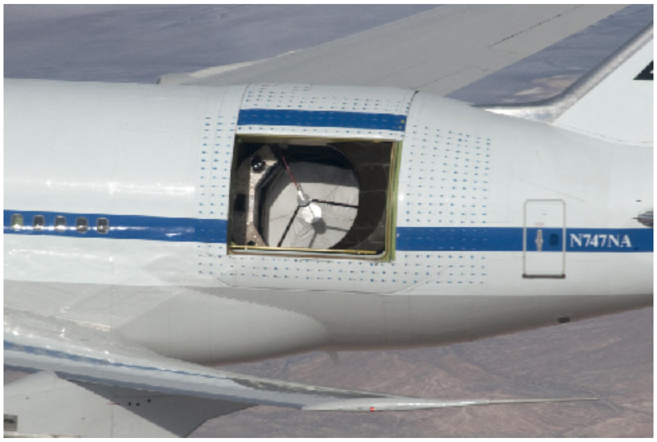
- List the advantages of making astronomical observations from space
- Explain the importance of the Hubble Space Telescope
- Describe some of the major space-based observatories astronomers use

Earth’s atmosphere blocks most radiation at wavelengths shorter than visible light, so we can only make direct ultraviolet, X-ray, and gamma ray observations from space (though indirect gamma ray observations can be made from Earth). Getting above the distorting effects of the atmosphere is also an advantage at visible and infrared wavelengths. The stars don’t “twinkle” in space, so the amount of detail you can observe is limited only by the size of your instrument. On the other hand, it is expensive to place telescopes into space, and repairs can present a major challenge. This is why astronomers continue to build telescopes for use on the ground as well as for launching into space.

Airborne and Space Infrared Telescopes

Water vapor, the main source of atmospheric interference for making infrared observations, is concentrated in the lower part of Earth’s atmosphere. For this reason, a gain of even a few hundred meters in elevation can make an important difference in the quality of an infared observatory site. Given the limitations of high mountains, most of which attract clouds and violent storms, and the fact that the ability of humans to perform complex tasks degrades at high altitudes, it was natural for astronomers to investigate the possibility of observing infrared waves from airplanes and ultimately from space.

Infrared observations from airplanes have been made since the 1960s, starting with a 15-centimeter telescope on board a Learjet. From 1974 through 1995, NASA operated a 0.9-meter airborne telescope flying regularly out of the Ames Research Center south of San Francisco. Observing from an altitude of 12 kilometers, the telescope was above 99% of the atmospheric water vapor. More recently, NASA (in partnership with the German Aerospace Center) has constructed a much larger 2.5-meter telescope, called the Stratospheric Observatory for Infrared Astronomy (SOFIA), which flies in a modified Boeing 747SP ([link]).
Stratospheric Observatory for Infrared Astronomy (SOFIA).

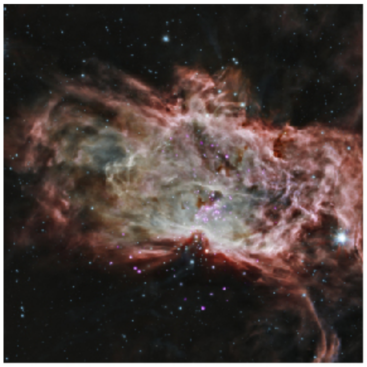


SOFIA allows observations to be made above most of Earth’s atmospheric water vapor. (credit: NASA)

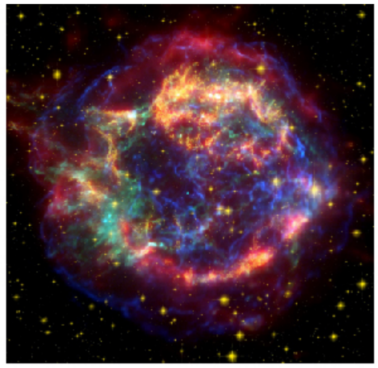
Note:
To find out more about SOFIA, watch this video provided by NASA’s Armstrong Flight Research Center.

Getting even higher and making observations from space itself have important advantages for infrared astronomy. First is the elimination of all interference from the atmosphere. Equally important is the opportunity to cool the entire optical system of the instrument in order to nearly eliminate infrared radiation from the telescope itself. If we tried to cool a telescope within the atmosphere, it would quickly become coated with condensing water vapor and other gases, making it useless. Only in the vacuum of space can optical elements be cooled to hundreds of degrees below freezing and still remain operational.

The first orbiting infrared observatory, launched in 1983, was the Infrared Astronomical Satellite (IRAS), built as a joint project by the United States, the Netherlands, and Britain. IRAS was equipped with a 0.6-meter telescope cooled to a temperature of less than 10 K. For the first time, the infrared sky could be seen as if it were night, rather than through a bright foreground of atmospheric and telescope emissions. IRAS carried out a rapid but comprehensive survey of the entire infrared sky over a 10-month period, cataloging about 350,000 sources of infrared radiation. Since then, several other infrared telescopes have operated in space with much better sensitivity and resolution due to improvements in infrared detectors. The most powerful of these infrared telescopes is the 0.85-meter Spitzer Space Telescope, which launched in 2003. A few of its observations are shown in [link]. With infrared observations, astronomers can detect cooler parts of cosmic objects, such as the dust clouds around star nurseries and the remnants of dying stars, that visible-light images don’t reveal.
Observations from the Spitzer Space Telescope (SST).



Flame nebula



Cassiopeia A

Helix nebula

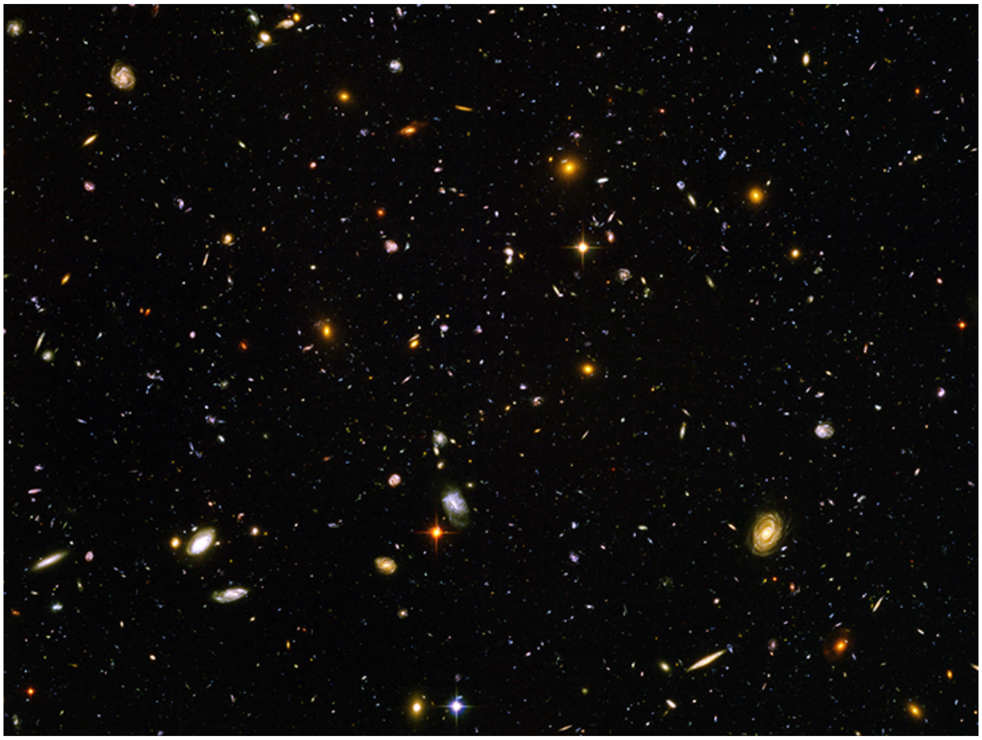
These infrared images—a region of star formation, the remnant of an exploded star, and a region where an old star is losing its outer shell—show just a few of the observations made and transmitted back to Earth from the SST. Since our eyes are not sensitive to infrared rays, we don't perceive colors from them. The colors in these images have been selected by astronomers to highlight details like the composition or temperature in these regions. (credit "Flame nebula": modification of work by NASA (X-ray: NASA/CXC/PSU/K.Getman, E.Feigelson, M.Kuhn & the MYStIX team; Infrared:NASA/JPL-Caltech); credit "Cassiopeia A": modification of work by NASA/JPL-Caltech; credit "Helix nebula": modification of work by NASA/JPL-Caltech)

Hubble Space Telescope

In April 1990, a great leap forward in astronomy was made with the launch of the Hubble Space Telescope (HST). With an aperture of 2.4 meters, this is the largest telescope put into space so far. (Its aperture was limited by the size of the payload bay in the Space Shuttle that served as its launch vehicle.) It was named for Edwin Hubble, the astronomer who discovered the expansion of the universe in the 1920s (whose work we will discuss in the chapters on Galaxies).

HST is operated jointly by NASA's Goddard Space Flight Center and the Space Telescope Science Institute in Baltimore. It was the first orbiting observatory designed to be serviced by Shuttle astronauts and, over the years since it was launched, they made several visits to improve or replace its initial instruments and to repair some of the systems that operate the spacecraft ([link])—though this repair program has now been discontinued, and no more visits or improvements will be made.

With the Hubble, astronomers have obtained some of the most detailed images of astronomical objects from the solar system outward to the most distant galaxies. Among its many great achievements is the Hubble Ultra-Deep Field, an image of a small region of the sky observed for almost 100 hours. It contains views of about 10,000 galaxies, some of which formed when the universe was just a few percent of its current age ([link]).
Hubble Ultra-Deep Field (HUDF).



The Hubble Space Telescope has provided an image of a specific region of space built from data collected between September 24, 2003, and January 16, 2004. These data allow us to search for galaxies that existed approximately 13 billion years ago. (credit: modification of work by NASA)

The HST’s mirror was ground and polished to a remarkable degree of accuracy. If we were to scale up its 2.4-meter mirror to the size of the entire continental United States, there would be no hill or valley larger than about 6 centimeters in its smooth surface. Unfortunately, after it was launched, scientists discovered that the primary mirror had a slight error in its *shape*, equal to roughly 1/50 the width of a human hair. Small as that sounds, it was enough to ensure that much of the light entering the telescope did not come to a clear focus and that all the images were blurry. (In a misplaced effort to save money, a complete test of the optical system had not been carried out before launch, so the error was not discovered until HST was in orbit.)

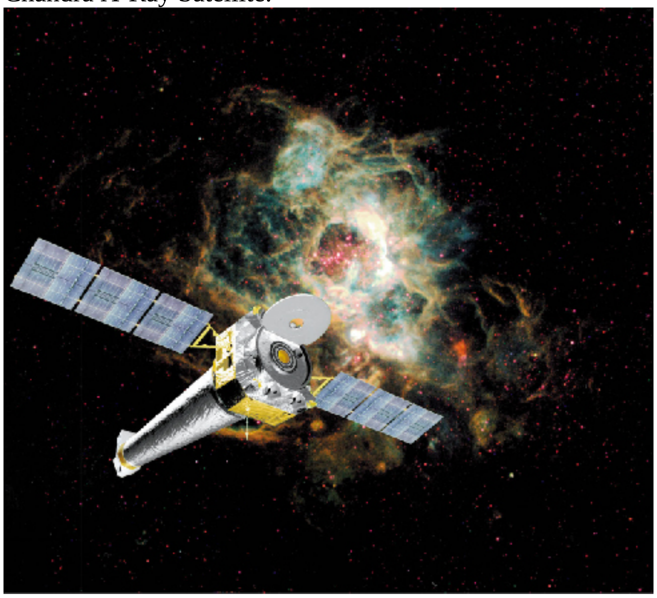
The solution was to do something very similar to what we do for astronomy students with blurry vision: put corrective optics in front of their eyes. In December 1993, in one of the most exciting and difficult space missions ever flown, astronauts captured the orbiting telescope and brought it back into the shuttle payload bay. There they installed a package containing compensating optics as well as a new, improved camera before releasing HST back into orbit. The telescope now works as it was intended to, and further missions to it were able to install even more advanced instruments to take advantage of its capabilities.

High-Energy Observatories

Ultraviolet, X-ray, and direct gamma-ray (high-energy electromagnetic wave) observations can be made only from space. Such observations first became possible in 1946, with V2 rockets captured from Germany after World War II. The US Naval Research Laboratory put instruments on these rockets for a series of pioneering flights, used initially to detect ultraviolet radiation from the Sun. Since then, many other rockets have been launched to make X-ray and ultraviolet observations of the Sun, and later of other celestial objects.

Beginning in the 1960s, a steady stream of high-energy observatories has been launched into orbit to reveal and explore the universe at short wavelengths. Among recent X-ray telescopes is the Chandra X-ray Observatory, which was launched in 1999 ([link]). It is producing X-ray images with unprecedented resolution and sensitivity. Designing instruments that can collect and focus energetic radiation like X-rays and gamma rays is an enormous technological challenge. The 2002 Nobel Prize in physics was awarded to Riccardo Giacconi, a pioneer in the field of building and launching sophisticated X-ray instruments. In 2008, NASA launched the Fermi Gamma-ray Space Telescope, designed to measure cosmic gamma rays at energies greater than any previous telescope, and thus able to collect radiation from some of the most energetic events in the universe.

Chandra X-Ray Satellite.



Chandra, the world's most powerful X-ray telescope, was developed by NASA and launched in July 1999. (credit: modification of work by NASA)

One major challenge is to design "mirrors" to reflect such penetrating radiation as X-rays and gamma rays, which normally pass straight through matter. However, although the technical details of design are more complicated, the three basic components of an observing system, as we explained earlier in this chapter, are the same at all wavelengths: a telescope to gather up the radiation, filters or instruments to sort the radiation according to wavelength, and some method of detecting and making a permanent record of the observations. [link] lists some of the most important active space observatories that humanity has launched.

Gamma-ray detections can also be made from Earth's surface by using the atmosphere as the primary detector. When a gamma ray hits our atmosphere, it accelerates charged particles (mostly electrons) in the atmosphere. Those energetic particles hit other particles in the atmosphere and give off their own radiation. The effect is a cascade of light and energy that can be detected on the ground. The VERITAS array in Arizona and the H.E.S.S. array in Namibia are two such ground-based gamma-ray observatories.

Recent Observatories in Space

Recent Observatories in Space

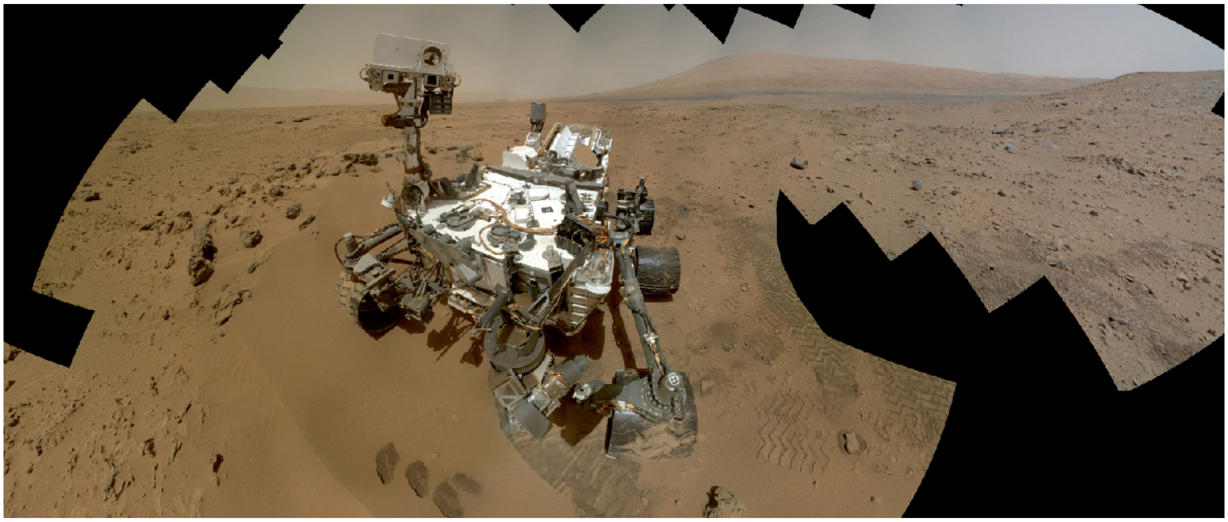
Observatory	Date Operation Began	Bands of the Spectrum	Notes	Website
Observatory	**Date Operation Began**	**Bands of the Spectrum**	**Notes**	**Website**
Hubble Space Telescope (HST)	1990	visible, UV, IR	2.4-m mirror; images and spectra	www.hubblesite.org
Chandra X-Ray Observatory	1999	X-rays	X-ray images and spectra	www.chandra.si.edu
XMM-Newton	1999	X-rays	X-ray spectroscopy	http://www.cosmos.esa.int/web/xmm-newton
International Gamma-Ray Astrophysics Laboratory (INTEGRAL)	2002	X- and gamma-rays	higher resolution gamma-ray images	http://sci.esa.int/integral/
Spitzer Space Telescope	2003	IR	0.85-m telescope	www.spitzer.caltech.edu
Fermi Gamma-ray Space Telescope	2008	gamma-rays	first high-energy gamma-ray observations	fermi.gsfc.nasa.gov
Kepler	2009	visible-light	planet finder	http://kepler.nasa.gov
Wide-field Infrared Survey Explorer (WISE)	2009	IR	whole-sky map, asteroid searches	www.nasa.gov/mission_pages/WISE/main
Gaia	2013	visible-light	Precise map of the Milky Way	http://sci.esa.int/gaia/
Transiting Exoplanet Survey Satellite (TESS)	2018	visible-light	Planet finder	http://tess.mit.edu

Infrared observations are made with telescopes aboard aircraft and in space, as well as from ground-based facilities on dry mountain peaks. Ultraviolet, X-ray, and gamma-ray observations must be made from above the atmosphere.

Many orbiting observatories have been flown to observe in these bands of the spectrum in the last few decades. The largest-aperture telescope in space is the Hubble Space telescope (HST), the most significant infrared telescope is Spitzer, and Chandra and Fermi are the premier X-ray and gamma-ray observatories, respectively.

Thinking Ahead
class="introduction"
“Self-Portrait” of Mars.

This picture was taken by the *Curiosity* Rover on Mars in 2012. The image is reconstructed digitally from 55 different images taken by a camera on the rover’s extended mast, so that the many positions of the mast (which acted like a selfie stick) are edited out. (credit: modification of work by NASA/JPL-Caltech/MSSS)



Surrounding the Sun is a complex system of worlds with a wide range of conditions: eight major planets, many dwarf planets, hundreds of moons, and countless smaller objects. Thanks largely to visits by spacecraft, we can now envision the members of the solar system as other worlds like our own, each with its own chemical and geological history, and unique sights that interplanetary tourists may someday visit. Some have called these past few decades the “golden age of planetary exploration,” comparable to the golden age of exploration in the fifteenth century, when great sailing ships plied Earth’s oceans and humanity became familiar with our own planet’s surface.

In this chapter, we discuss our planetary system and introduce the idea of comparative planetology—studying how the planets work by comparing them with one another. We want to get to know the planets not only for what we can learn about them, but also to see what they can tell us about the origin and evolution of the entire solar system. In the upcoming chapters, we describe the better-known members of the solar system and begin to compare them to the thousands of planets that have been discovered recently, orbiting other stars.

Overview of Our Planetary System

LEARNING OBJECTIVES

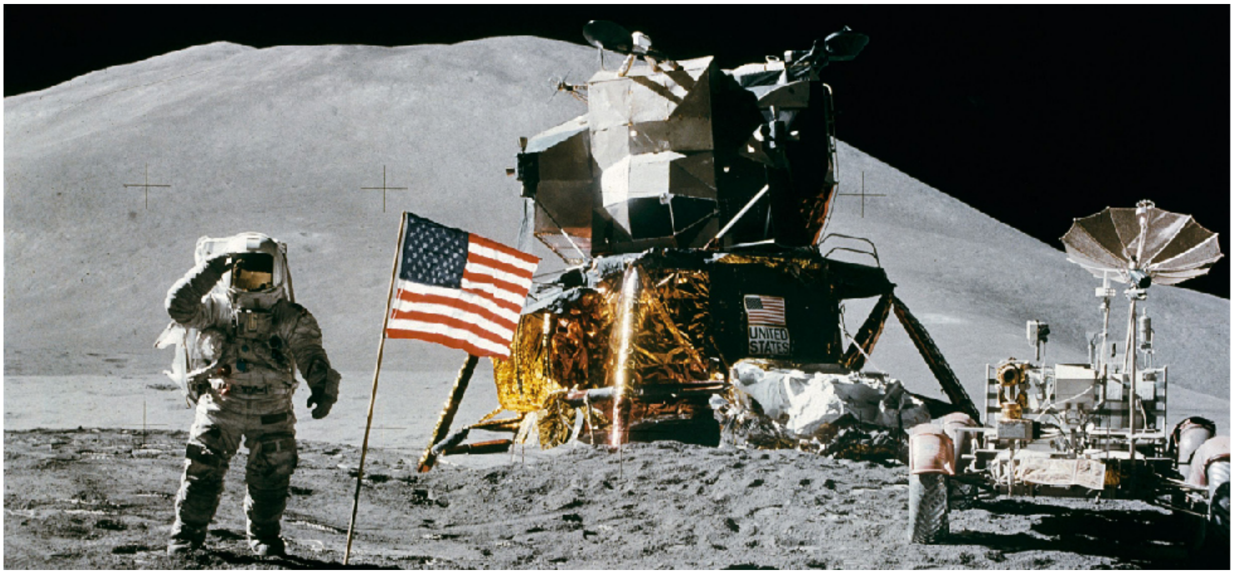
By the end of this section, you will be able to:

- Describe how the objects in our solar system are identified, explored, and characterized
- Describe the types of small bodies in our solar system, their locations, and how they formed
- Model the solar system with distances from everyday life to better comprehend distances in space

The solar system[footnote] consists of the Sun and many smaller objects: the planets, their moons and rings, and such "debris" as asteroids, comets, and dust. Decades of observation and spacecraft exploration have revealed that most of these objects formed together with the Sun about 4.5 billion years ago. They represent clumps of material that condensed from an enormous cloud of gas and dust. The central part of this cloud became the Sun, and a small fraction of the material in the outer parts eventually formed the other objects.
The generic term for a group of planets and other bodies circling a star is *planetary system*. Ours is called the *solar system* because our Sun is sometimes called *Sol*. Strictly speaking, then, there is only one solar system; planets orbiting other stars are in planetary systems.

During the past 50 years, we have learned more about the solar system than anyone imagined before the space age. In addition to gathering information with powerful new telescopes, we have sent spacecraft directly to many members of the planetary system. (Planetary astronomy is the only branch of our science in which we can, at least vicariously, travel to the objects we want to study.) With evocative names such as Voyager, Pioneer, *Curiosity*, and Pathfinder, our robot explorers have flown past, orbited, or landed on every planet, returning images and data that have dazzled both astronomers and the public. In the process, we have also investigated two dwarf planets, hundreds of fascinating moons, four ring systems, a dozen asteroids, and several comets (smaller members of our solar system that we will discuss later).

Our probes have penetrated the atmosphere of Jupiter and landed on the surfaces of Venus, Mars, our Moon, Saturn's moon Titan, the asteroids Eros and Itokawa, and the Comet Churyumov-Gerasimenko (usually referred to as 67P). Humans have set foot on the Moon and returned samples of its surface soil for laboratory analysis ([link]). We have even discovered other places in our solar system that might be able to support some kind of life.
Astronauts on the Moon.



The lunar lander and surface rover from the Apollo 15 mission are seen in this view of the one place beyond Earth that has been explored directly by humans. (credit: modification of work by David R. Scott, NASA)

Note:
View this gallery of NASA images that trace the history of the Apollo mission.

An Inventory

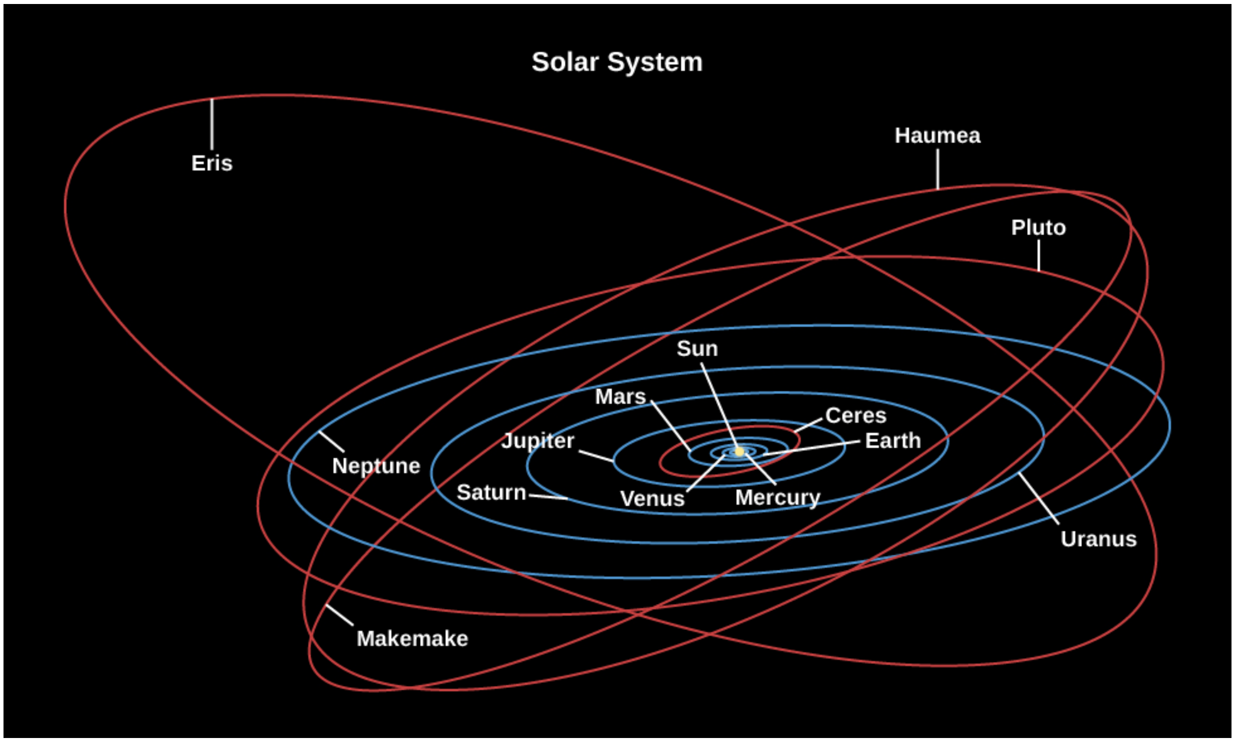
The Sun, a star that is brighter than about 80% of the stars in the Galaxy, is by far the most massive member of the solar system, as shown in [link]. It is an enormous ball about 1.4 million kilometers in diameter, with surface layers of incandescent gas and an interior temperature of millions of degrees. The Sun will be discussed in later chapters as our first, and best-studied, example of a star.

Mass of Members of the Solar System	
Object	**Percentage of Total Mass of Solar System**
Sun	99.80
Jupiter	0.10

Mass of Members of the Solar System	
Object	**Percentage of Total Mass of Solar System**
Comets	0.0005–0.03 (estimate)
All other planets and dwarf planets	0.04
Moons and rings	0.00005
Asteroids	0.000002 (estimate)
Cosmic dust	0.0000001 (estimate)

[link] also shows that most of the material of the planets is actually concentrated in the largest one, Jupiter, which is more massive than all the rest of the planets combined. Astronomers were able to determine the masses of the planets centuries ago using Kepler's laws of planetary motion and Newton's law of gravity to measure the planets' gravitational effects on one another or on moons that orbit them (see Orbits and Gravity). Today, we make even more precise measurements of their masses by tracking their gravitational effects on the motion of spacecraft that pass near them.

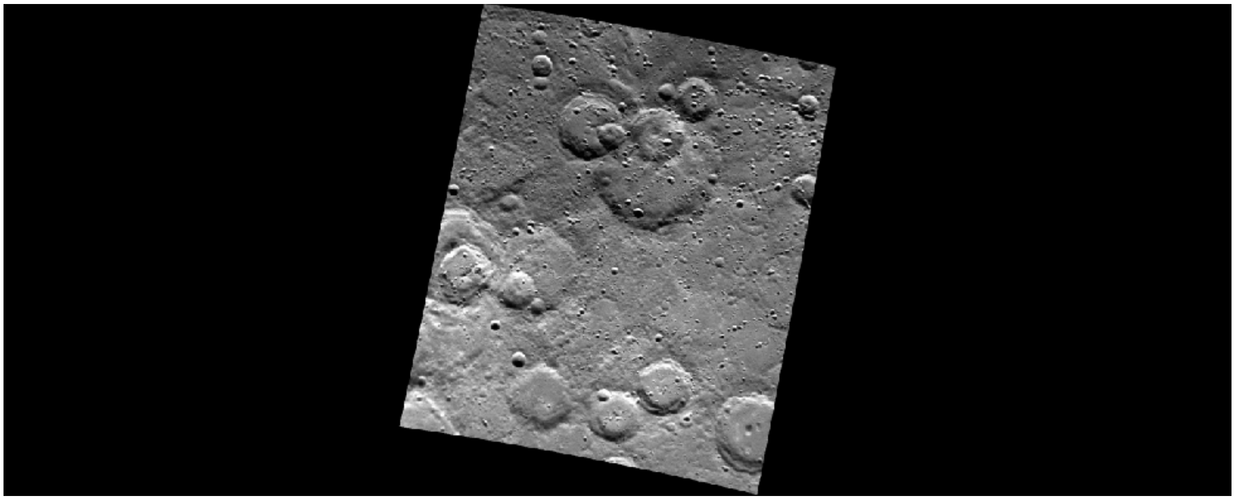
Beside Earth, five other planets were known to the ancients—Mercury, Venus, Mars, Jupiter, and Saturn—and two were discovered after the invention of the telescope: Uranus and Neptune. The eight planets all revolve in the same direction around the Sun. They orbit in approximately the same plane, like cars traveling on concentric tracks on a giant, flat racecourse. Each planet stays in its own "traffic lane," following a nearly circular orbit about the Sun and obeying the "traffic" laws discovered by Galileo, Kepler, and Newton. Besides these planets, we have also been discovering smaller worlds beyond Neptune that are called trans-Neptunian objects or TNOs (see [link]). The first to be found, in 1930, was Pluto, but others have been discovered during the twenty-first century. One of them, Eris, is about the same size as Pluto and has at least one moon (Pluto has five known moons.) The largest TNOs are also classed as *dwarf planets*, as is the largest asteroid, Ceres. (Dwarf planets will be discussed further in the chapter on Rings, Moons, and Pluto). To date, more than 1750 of these TNOs have been discovered.
Orbits of the Planets.



All eight major planets orbit the Sun in roughly the same plane. The five currently known dwarf planets are also shown: Eris, Haumea, Pluto, Ceres, and Makemake. Note that Pluto's orbit is not in the plane of the planets.

Each of the planets and dwarf planets also rotates (spins) about an axis running through it, and in most cases the direction of rotation is the same as the direction of revolution about the Sun. The exceptions are Venus, which rotates backward very slowly (that is, in a retrograde direction), and Uranus and Pluto, which also have strange rotations, each spinning about an axis tipped nearly on its side. We do not yet know the spin orientations of Eris, Haumea, and Makemake.

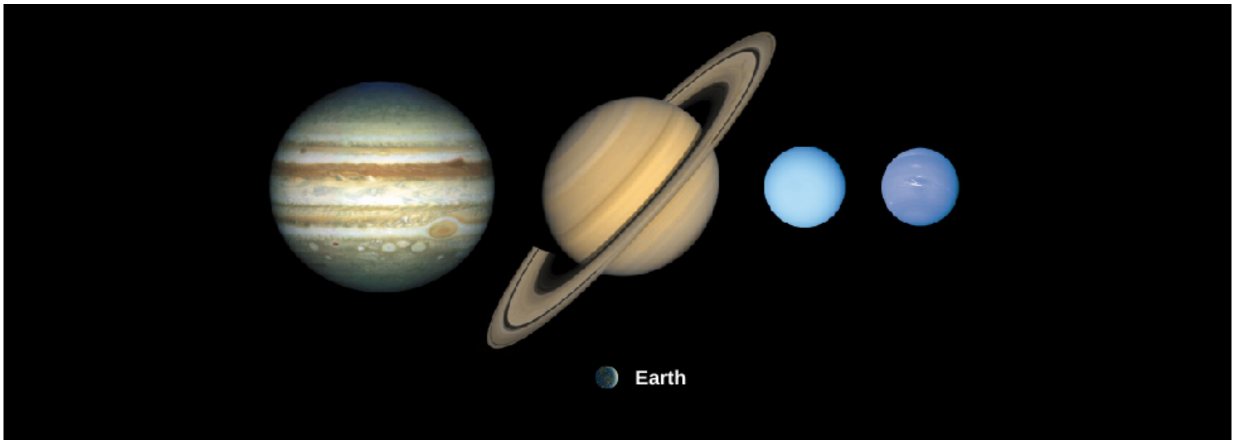
The four planets closest to the Sun (Mercury through Mars) are called the inner or **terrestrial planets**. Often, the Moon is also discussed as a part of this group, bringing the total of terrestrial objects to five. (We generally call Earth's satellite "the Moon," with a capital M, and the other satellites "moons," with lowercase m's.) The terrestrial planets are relatively small worlds, composed primarily of rock and metal. All of them have solid surfaces that bear the records of their geological history in the forms of craters, mountains, and volcanoes ([link]).
Surface of Mercury.



The pockmarked face of the terrestrial world of Mercury is more typical of the inner planets than the watery surface of Earth. This black-and-white image, taken with the Mariner 10 spacecraft, shows a region more than 400 kilometers wide. (credit: modification of work by NASA/John Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington)

The next four planets (Jupiter through Neptune) are much larger and are composed primarily of lighter ices, liquids, and gases. We call these four the jovian planets (after “Jove,” another name for Jupiter in mythology) or **giant planets**—a name they richly deserve ([link]). About 1,300 Earths could fit inside Jupiter, for example. These planets do not have solid surfaces on which future explorers might land. They are more like vast, spherical oceans with much smaller, dense cores.

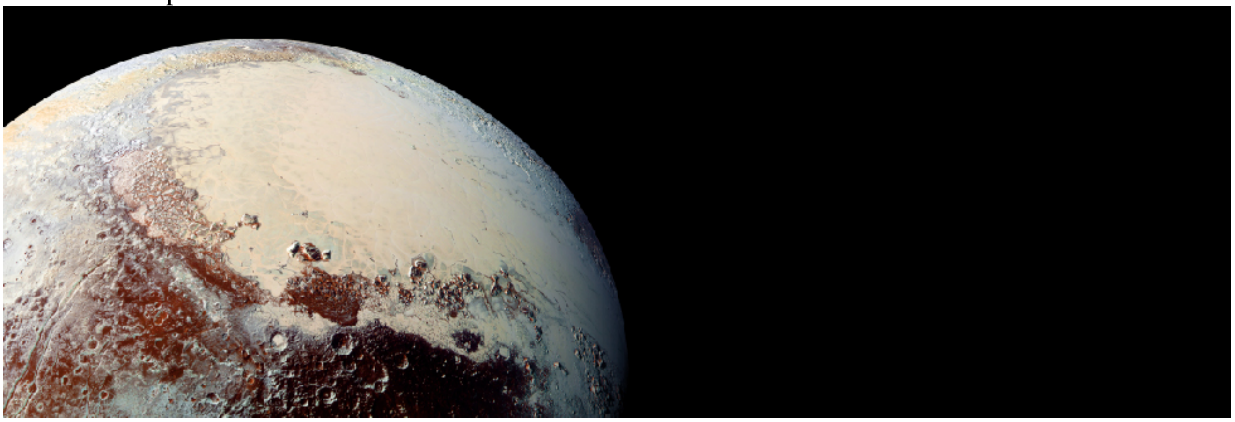
The Four Giant Planets.



This montage shows the four giant planets: Jupiter, Saturn, Uranus, and Neptune. Below them, Earth is shown to scale. (credit: modification of work by NASA, Solar System Exploration)

Near the outer edge of the system lies Pluto, which was the first of the distant icy worlds to be discovered beyond Neptune (Pluto was visited by a spacecraft, the NASA New Horizons mission, in 2015 [see [link]]). [link] summarizes some of the main facts about the planets.

Pluto Close-up.



This intriguing image from the New Horizons spacecraft, taken when it flew by the dwarf planet in July 2015, shows some of its complex surface features. The rounded white area is called the Sputnik Plain, after humanity's first spacecraft. (credit: modification of work by NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

The Planets

Name	Distance from Sun (AU)[footnote] An AU (or astronomical unit) is the distance from Earth to the Sun.	Revolution Period (y)	Diameter (km)	Mass (10^{23} kg)	Density (g/cm^3)[footnote] We give densities in units where the density of water is 1 g/cm^3. To get densities in units of kg/m^3, multiply the given value by 1000.
Mercury	0.39	0.24	4,878	3.3	5.4
Venus	0.72	0.62	12,120	48.7	5.2
Earth	1.00	1.00	12,756	59.8	5.5
Mars	1.52	1.88	6,787	6.4	3.9
Jupiter	5.20	11.86	142,984	18,991	1.3

Saturn	9.54	29.46	120,536	5686	0.7
Uranus	19.18	84.07	51,118	866	1.3
Neptune	30.06	164.82	49,660	1030	1.6

Example:
Comparing Densities
Let's compare the densities of several members of the solar system. The density of an object equals its mass divided by its volume. The volume (V) of a sphere (like a planet) is calculated using the equation
Equation:

$$V = \frac{4}{3}\pi R^3$$

where π (the Greek letter pi) has a value of approximately 3.14. Although planets are not perfect spheres, this equation works well enough. The masses and diameters of the planets are given in [link]. For data on selected moons, see Appendix G. Let's use Saturn's moon Mimas as our example, with a mass of 4×10^{19} kg and a diameter of approximately 400 km (radius, 200 km = 2×10^5 m).
Solution
The volume of Mimas is
Equation:

$$\frac{4}{3} \times 3.14 \times \left(2 \times 10^5 \text{ m}\right)^3 = 3.3 \times 10^{16} \text{ m}^3.$$

Density is mass divided by volume:
Equation:

$$\frac{4 \times 10^{19} \text{ kg}}{3.3 \times 10^{16} \text{ m}^3} = 1.2 \times 10^3 \text{ kg/m}^3.$$

Note that the density of water in these units is 1000 kg/m^3, so Mimas must be made mainly of ice, not rock. (Note that the density of Mimas given in Appendix G is 1.2, but the units used there are different. In that table, we give density in units of
g/cm^3, for which the density of water equals 1. Can you show, by converting units, that 1 g/cm^3 is the same as 1000 kg/m^3?)
Check Your Learning
Calculate the average density of our own planet, Earth. Show your work. How does it compare to the density of an ice moon like Mimas? See [link] for data.

Note:
Answer:

For a sphere,

$$\text{density} = \frac{\text{mass}}{\left(\frac{4}{3}\pi R^3\right)}\ \text{kg/m}^3.$$

For Earth, then,

$$\text{density} = \frac{6 \times 10^{24}\,\text{kg}}{4.2 \times 2.6 \times 10^{20}\,\text{m}^3} = 5.5\ \times\ 10^3\,\text{kg/m}^3.$$

This density is four to five times greater than Mimas'. In fact, Earth is the densest of the planets.

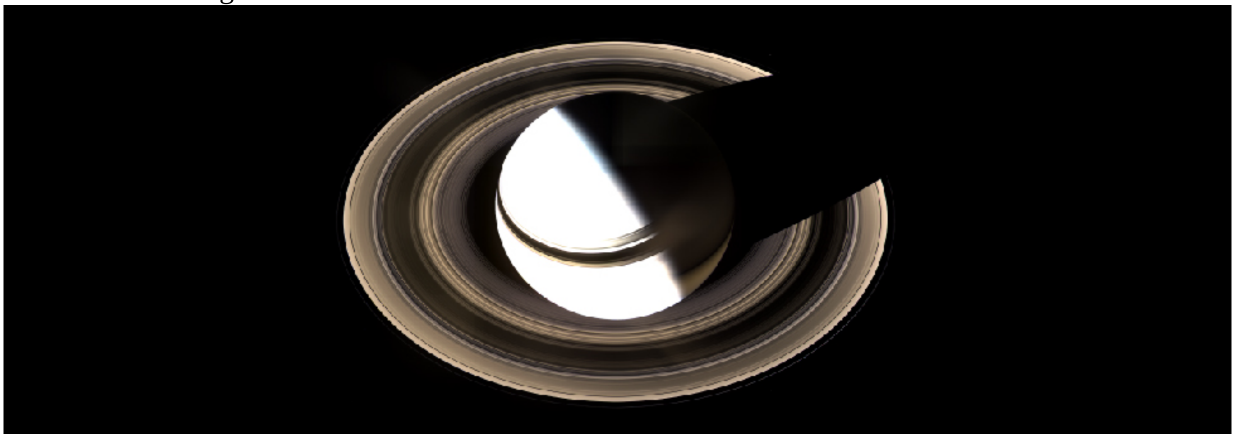
Note:
Learn more about NASA's mission to Pluto and see high-resolution images of Pluto's moon Charon.

Smaller Members of the Solar System

Most of the planets are accompanied by one or more moons; only Mercury and Venus move through space alone. There are more than 180 known moons orbiting planets and dwarf planets (see Appendix G for a listing of the larger ones), and undoubtedly many other small ones remain undiscovered. The largest of the moons are as big as small planets and just as interesting. In addition to our Moon, they include the four largest moons of Jupiter (called the Galilean moons, after their discoverer) and the largest moons of Saturn and Neptune (confusingly named Titan and Triton).

Each of the giant planets also has rings made up of countless small bodies ranging in size from mountains to mere grains of dust, all in orbit about the equator of the planet. The bright rings of Saturn are, by far, the easiest to see. They are among the most beautiful sights in the solar system ([link]). But, all four ring systems are interesting to scientists because of their complicated forms, influenced by the pull of the moons that also orbit these giant planets.

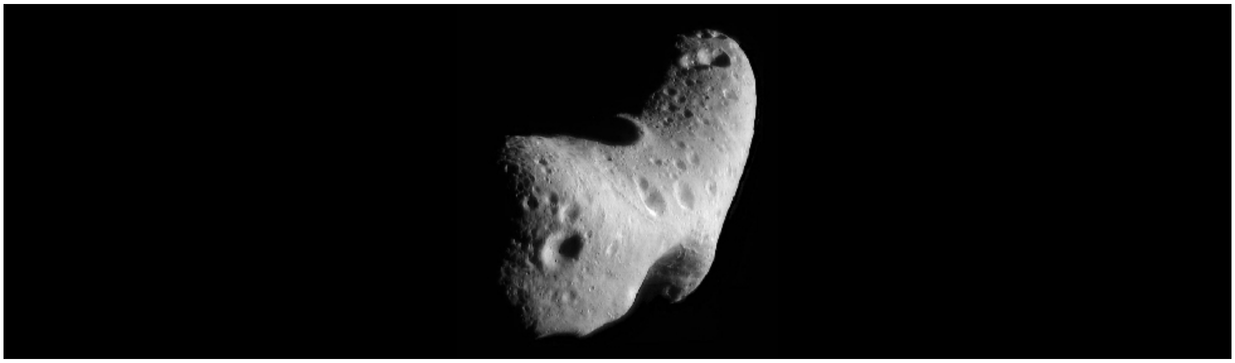
Saturn and Its Rings.



This 2007 Cassini image shows Saturn and its complex system of rings, taken from a distance of about 1.2 million kilometers. This natural-color image is a composite of 36 images taken over the course of 2.5 hours. (credit: modification of work by NASA/JPL/Space Science Institute)

The solar system has many other less-conspicuous members. Another group is the **asteroids**, rocky bodies that orbit the Sun like miniature planets, mostly in the space between Mars and Jupiter (although some do cross the orbits of planets like Earth—see [link]). Most asteroids are remnants of the initial population of the solar system that existed before the planets themselves formed. Some of the smallest moons of the planets, such as the moons of Mars, are very likely captured asteroids.

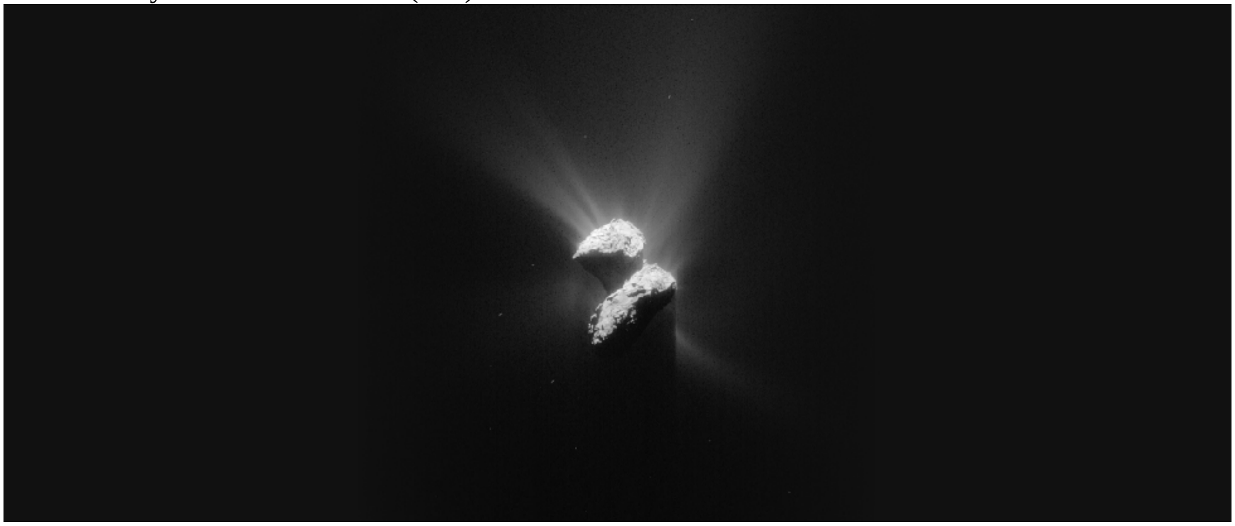
Asteroid Eros.



This small Earth-crossing asteroid image was taken by the NEAR-Shoemaker spacecraft from an altitude of about 100 kilometers. This view of the heavily cratered surface is about 10 kilometers wide. The spacecraft orbited Eros for a year before landing gently on its surface. (credit: modification of work by NASA/JHUAPL)

Another class of small bodies is composed mostly of ice, made of frozen gases such as water, carbon dioxide, and carbon monoxide; these objects are called **comets** (see [link]). Comets also are remnants from the formation of the solar system, but they were formed and continue (with rare exceptions) to orbit the Sun in distant, cooler regions—stored in a sort of cosmic deep freeze. This is also the realm of the larger icy worlds, called dwarf planets.

Comet Churyumov-Gerasimenko (67P).



This image shows Comet Churyumov-Gerasimenko, also known as 67P, near its closest approach to the Sun in 2015, as seen from the *Rosetta* spacecraft. Note the jets of gas escaping from the

solid surface. (credit: modification of work by ESA/Rosetta/NAVACAM, CC BY-SA IGO 3.0)

Finally, there are countless grains of broken rock, which we call cosmic dust, scattered throughout the solar system. When these particles enter Earth’s atmosphere (as millions do each day) they burn up, producing a brief flash of light in the night sky known as a **meteor** (meteors are often referred to as shooting stars). Occasionally, some larger chunk of rocky or metallic material survives its passage through the atmosphere and lands on Earth. Any piece that strikes the ground is known as a **meteorite**. (You can see meteorites on display in many natural history museums and can sometimes even purchase pieces of them from gem and mineral dealers.)

Note:
Carl Sagan: Solar System Advocate
The best-known astronomer in the world during the 1970s and 1980s, Carl Sagan devoted most of his professional career to studying the planets and considerable energy to raising public awareness of what we can learn from exploring the solar system (see [link]). Born in Brooklyn, New York, in 1934, Sagan became interested in astronomy as a youngster; he also credits science fiction stories for sustaining his fascination with what’s out in the universe.
Carl Sagan (1934–1996) and Neil deGrasse Tyson.

Sagan was Tyson’s inspiration to become a scientist. (credit “Sagan”: modification of work by NASA, JPL; credit “Tyson”: modification of work by Bruce F. Press)

In the early 1960s, when many scientists still thought Venus might turn out to be a hospitable place, Sagan calculated that the thick atmosphere of Venus could act like a giant greenhouse, keeping the heat in and raising the temperature enormously. He showed that the seasonal changes astronomers had seen on Mars were caused, not by vegetation, but by wind-blown dust. He was a member of the

scientific teams for many of the robotic missions that explored the solar system and was instrumental in getting NASA to put a message-bearing plaque aboard the Pioneer spacecraft, as well as audio-video records on the Voyager spacecraft—all of them destined to leave our solar system entirely and send these little bits of Earth technology out among the stars.

To encourage public interest and public support of planetary exploration, Sagan helped found The Planetary Society, now the largest space-interest organization in the world. He was a tireless and eloquent advocate of the need to study the solar system close-up and the value of learning about other worlds in order to take better care of our own.

Sagan simulated conditions on early Earth to demonstrate how some of life's fundamental building blocks might have formed from the "primordial soup" of natural compounds on our planet. In addition, he and his colleagues developed computer models showing the consequences of nuclear war for Earth would be even more devastating than anyone had thought (this is now called the nuclear winter hypothesis) and demonstrating some of the serious consequences of continued pollution of our atmosphere.

Sagan was perhaps best known, however, as a brilliant popularizer of astronomy and the author of many books on science, including the best-selling *Cosmos*, and several evocative tributes to solar system exploration such as *The Cosmic Connection* and *Pale Blue Dot*. His book *The Demon Haunted World*, completed just before his death in 1996, is perhaps the best antidote to fuzzy thinking about pseudo-science and irrationality in print today. An intriguing science fiction novel he wrote, titled *Contact*, which became a successful film as well, is still recommended by many science instructors as a scenario for making contact with life elsewhere that is much more reasonable than most science fiction.

Sagan was a master, too, of the television medium. His 13-part public television series, *Cosmos*, was seen by an estimated 500 million people in 60 countries and has become one of the most-watched series in the history of public broadcasting. A few astronomers scoffed at a scientist who spent so much time in the public eye, but it is probably fair to say that Sagan's enthusiasm and skill as an explainer won more friends for the science of astronomy than anyone or anything else in the second half of the twentieth century.

In the two decades since Sagan's death, no other scientist has achieved the same level of public recognition. Perhaps closest is the director of the Hayden Planetarium, Neil deGrasse Tyson, who followed in Sagan's footsteps by making an updated version of the *Cosmos* program in 2014. Tyson is quick to point out that Sagan was his inspiration to become a scientist, telling how Sagan invited him to visit for a day at Cornell when he was a high school student looking for a career. However, the media environment has fragmented a great deal since Sagan's time. It is interesting to speculate whether Sagan could have adapted his communication style to the world of cable television, Twitter, Facebook, and podcasts.

Note:

Two imaginative videos provide a tour of the solar system objects we have been discussing. Shane Gellert's I Need Some Space uses NASA photography and models to show the various worlds with which we share our system. In the more science fiction-oriented Wanderers video, we see some of the planets and moons as tourist destinations for future explorers, with commentary taken from recordings by Carl Sagan.

A Scale Model of the Solar System

Astronomy often deals with dimensions and distances that far exceed our ordinary experience. What does 1.4 billion kilometers—the distance from the Sun to Saturn—really mean to anyone? It can be helpful to visualize such large systems in terms of a scale model.

In our imaginations, let us build a scale model of the solar system, adopting a scale factor of 1 billion (10^9)—that is, reducing the actual solar system by dividing every dimension by a factor of 10^9. Earth, then, has a diameter of 1.3 centimeters, about the size of a grape. The Moon is a pea orbiting this at a distance of 40 centimeters, or a little more than a foot away. The Earth-Moon system fits into a standard backpack.

In this model, the Sun is nearly 1.5 meters in diameter, about the average height of an adult, and our Earth is at a distance of 150 meters—about one city block—from the Sun. Jupiter is five blocks away from the Sun, and its diameter is 15 centimeters, about the size of a very large grapefruit. Saturn is 10 blocks from the Sun; Uranus, 20 blocks; and Neptune, 30 blocks. Pluto, with a distance that varies quite a bit during its 249-year orbit, is currently just beyond 30 blocks and getting farther with time. Most of the moons of the outer solar system are the sizes of various kinds of seeds orbiting the grapefruit, oranges, and lemons that represent the outer planets.

In our scale model, a human is reduced to the dimensions of a single atom, and cars and spacecraft to the size of molecules. Sending the Voyager spacecraft to Neptune involves navigating a single molecule from the Earth–grape toward a lemon 5 kilometers away with an accuracy equivalent to the width of a thread in a spider's web.

If that model represents the solar system, where would the nearest stars be? If we keep the same scale, the closest stars would be tens of thousands of kilometers away. If you built this scale model in the city where you live, you would have to place the representations of these stars on the other side of Earth or beyond.

By the way, model solar systems like the one we just presented have been built in cities throughout the world. In Sweden, for example, Stockholm's huge Globe Arena has become a model for the Sun, and Pluto is represented by a 12-centimeter sculpture in the small town of Delsbo, 300 kilometers away. Another model solar system is in Washington on the Mall between the White House and Congress (perhaps proving they are worlds apart?).

Note:
Names in the Solar System
We humans just don't feel comfortable until something has a name. Types of butterflies, new elements, and the mountains of Venus all need names for us to feel we are acquainted with them. How do we give names to objects and features in the solar system?
Planets and moons are named after gods and heroes in Greek and Roman mythology (with a few exceptions among the moons of Uranus, which have names drawn from English literature). When William Herschel, a German immigrant to England, first discovered the planet we now call Uranus, he wanted to name it Georgium Sidus (George's star) after King George III of his adopted country. This caused such an outcry among astronomers in other nations, however, that the classic tradition was upheld—and has been maintained ever since. Luckily, there were a lot of minor gods in the ancient pantheon, so plenty of names are left for the many small moons we are discovering around the giant planets. (Appendix G lists the larger moons).
Comets are often named after their discoverers (offering an extra incentive to comet hunters). Asteroids are named by their discoverers after just about anyone or anything they want. Recently,

asteroid names have been used to recognize people who have made significant contributions to astronomy, including the three original authors of this book.

That was pretty much all the naming that was needed while our study of the solar system was confined to Earth. But now, our spacecraft have surveyed and photographed many worlds in great detail, and each world has a host of features that also need names. To make sure that naming things in space remains multinational, rational, and somewhat dignified, astronomers have given the responsibility of approving names to a special committee of the International Astronomical Union (IAU), the body that includes scientists from every country that does astronomy.

This IAU committee has developed a set of rules for naming features on other worlds. For example, craters on Venus are named for women who have made significant contributions to human knowledge and welfare. Volcanic features on Jupiter’s moon Io, which is in a constant state of volcanic activity, are named after gods of fire and thunder from the mythologies of many cultures. Craters on Mercury commemorate famous novelists, playwrights, artists, and composers. On Saturn’s moon Tethys, all the features are named after characters and places in Homer’s great epic poem, *The Odyssey*. As we explore further, it may well turn out that more places in the solar system need names than Earth history can provide. Perhaps by then, explorers and settlers on these worlds will be ready to develop their own names for the places they may (if but for a while) call home.

You may be surprised to know that the meaning of the word *planet* has recently become controversial because we have discovered many other planetary systems that don’t look very much like our own. Even within our solar system, the planets differ greatly in size and chemical properties. The biggest dispute concerns Pluto, which is much smaller than the other eight major planets. The category of dwarf planet was invented to include Pluto and similar icy objects beyond Neptune. But is a dwarf planet also a planet? Logically, it should be, but even this simple issue of grammar has been the subject of heated debate among both astronomers and the general public.

Key Concepts and Summary

Our solar system currently consists of the Sun, eight planets, five dwarf planets, nearly 200 known moons, and a host of smaller objects. The planets can be divided into two groups: the inner terrestrial planets and the outer giant planets. Pluto, Eris, Haumea, and Makemake do not fit into either category; as icy dwarf planets, they exist in an ice realm on the fringes of the main planetary system. The giant planets are composed mostly of liquids and gases. Smaller members of the solar system include asteroids (including the dwarf planet Ceres), which are rocky and metallic objects found mostly between Mars and Jupiter; comets, which are made mostly of frozen gases and generally orbit far from the Sun; and countless smaller grains of cosmic dust. When a meteor survives its passage through our atmosphere and falls to Earth, we call it a meteorite.

Glossary

asteroid
: a stony or metallic object orbiting the Sun that is smaller than a major planet but that shows no evidence of an atmosphere or of other types of activity associated with comets

comet
: a small body of icy and dusty matter that revolves about the Sun; when a comet comes near the Sun, some of its material vaporizes, forming a large head of tenuous gas and often a tail

giant planet

any of the planets Jupiter, Saturn, Uranus, and Neptune in our solar system, or planets of roughly that mass and composition in other planetary systems

meteor

a small piece of solid matter that enters Earth's atmosphere and burns up, popularly called a *shooting star* because it is seen as a small flash of light

meteorite

a portion of a meteor that survives passage through an atmosphere and strikes the ground

terrestrial planet

any of the planets Mercury, Venus, Earth, or Mars; sometimes the Moon is included in the list

Composition and Structure of Planets

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the characteristics of the giant planets, terrestrial planets, and small bodies in the solar system
- Explain what influences the temperature of a planet's surface
- Explain why there is geological activity on some planets and not on others

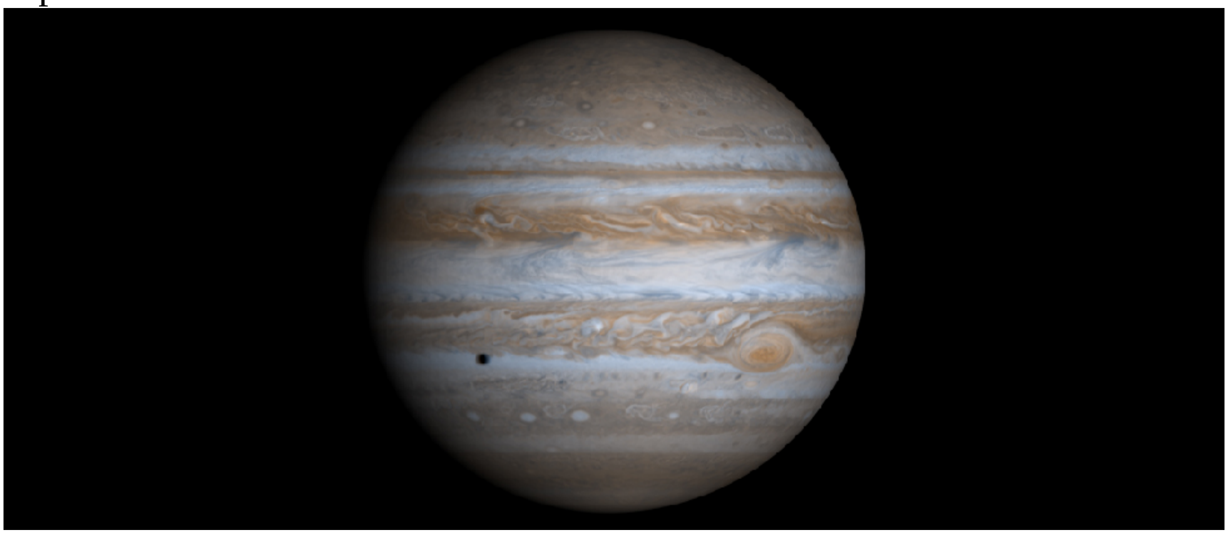
The fact that there are two distinct kinds of planets—the rocky terrestrial planets and the gas-rich jovian planets—leads us to believe that they formed under different conditions. Certainly their compositions are dominated by different elements. Let us look at each type in more detail.

The Giant Planets

The two largest planets, Jupiter and Saturn, have nearly the same chemical makeup as the Sun; they are composed primarily of the two elements hydrogen and helium, with 75% of their mass being hydrogen and 25% helium. On Earth, both hydrogen and helium are gases, so Jupiter and Saturn are sometimes called gas planets. But, this name is misleading. Jupiter and Saturn are so large that the gas is compressed in their interior until the hydrogen becomes a liquid. Because the bulk of both planets consists of compressed, liquefied hydrogen, we should really call them liquid planets.

Under the force of gravity, the heavier elements sink toward the inner parts of a liquid or gaseous planet. Both Jupiter and Saturn, therefore, have cores composed of heavier rock, metal, and ice, but we cannot see these regions directly. In fact, when we look down from above, all we see is the atmosphere with its swirling clouds ([link]). We must infer the existence of the denser core inside these planets from studies of each planet’s gravity.

Jupiter.



This true-color image of Jupiter was taken from the Cassini spacecraft in 2000. (credit: modification of work by NASA/JPL/University of Arizona)

Uranus and Neptune are much smaller than Jupiter and Saturn, but each also has a core of rock, metal, and ice. Uranus and Neptune were less efficient at attracting hydrogen and helium gas, so they have much smaller atmospheres in proportion to their cores.

Chemically, each giant planet is dominated by hydrogen and its many compounds. Nearly all the oxygen present is combined chemically with hydrogen to form water (H_2O). Chemists call such a hydrogen-dominated composition *reduced*. Throughout the outer solar system, we find abundant water (mostly in the form of ice) and reducing chemistry.

The Terrestrial Planets

The terrestrial planets are quite different from the giants. In addition to being much smaller, they are composed primarily of rocks and metals. These, in turn, are made of elements that are less common in the universe as a whole. The most abundant rocks, called silicates, are made of silicon and oxygen, and the most common metal is iron. We can tell from their densities (see [link]) that Mercury has the greatest proportion of metals (which are denser) and the Moon has the lowest. Earth, Venus, and Mars all have roughly similar bulk compositions: about one third of their mass consists of iron-nickel or iron-sulfur combinations; two thirds is made of silicates. Because these planets are largely composed of oxygen compounds (such as the silicate minerals of their crusts), their chemistry is said to be *oxidized*.

When we look at the internal structure of each of the terrestrial planets, we find that the densest metals are in a central core, with the lighter silicates near the surface. If these planets were liquid, like the giant planets, we could understand this effect as the result the sinking of heavier elements due to the pull of gravity. This leads us to conclude that, although the terrestrial planets are solid today, at one time they must have been hot enough to melt.

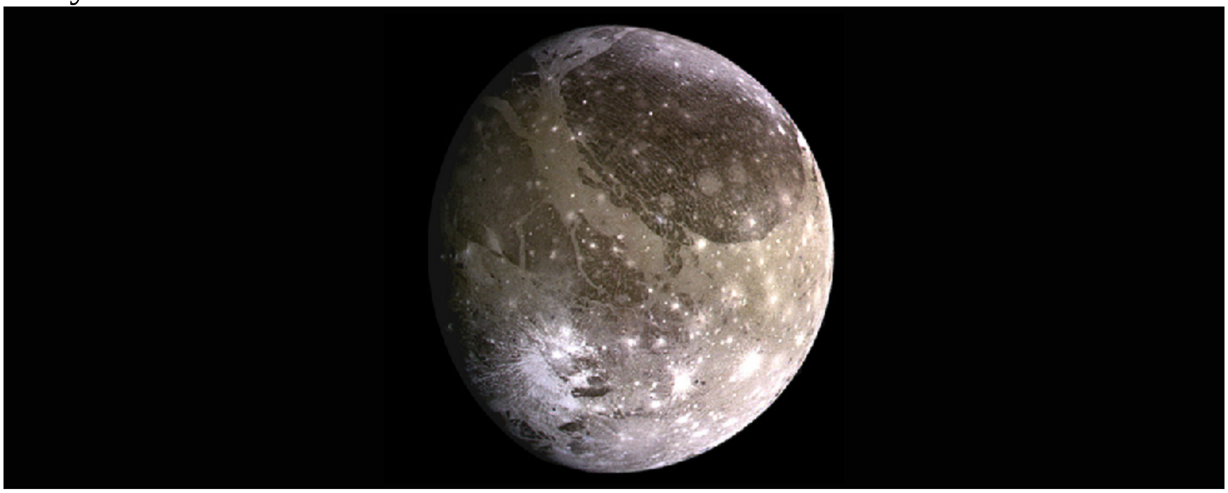
Differentiation is the process by which gravity helps separate a planet's interior into layers of different compositions and densities. The heavier metals sink to form a core, while the lightest minerals float to the surface to form a crust. Later, when the planet cools, this layered structure is preserved. In order for a rocky planet to differentiate, it must be heated to the melting point of rocks, which is typically more than 1300 K.

Moons, Asteroids, and Comets

Chemically and structurally, Earth's Moon is like the terrestrial planets, but most moons are in the outer solar system, and they have compositions similar to the cores of the giant planets around which they orbit. The three largest moons—Ganymede and Callisto in the jovian system, and Titan in the saturnian system—are composed half of frozen water, and half of rocks and metals. Most of these moons differentiated during formation, and today

they have cores of rock and metal, with upper layers and crusts of very cold and—thus very hard—ice ([link]).

Ganymede.



This view of Jupiter’s moon Ganymede was taken in June 1996 by the Galileo spacecraft. The brownish gray color of the surface indicates a dusty mixture of rocky material and ice. The bright spots are places where recent impacts have uncovered fresh ice from underneath. (credit: modification of work by NASA/JPL)

Most of the asteroids and comets, as well as the smallest moons, were probably never heated to the melting point. However, some of the largest asteroids, such as Vesta, appear to be differentiated; others are fragments from differentiated bodies. Because most asteroids and comets retain their original composition, they represent relatively unmodified material dating back to the time of the formation of the solar system. In a sense, they act as chemical fossils, helping us to learn about a time long ago whose traces have been erased on larger worlds.

Temperatures: Going to Extremes

Generally speaking, the farther a planet or moon is from the Sun, the cooler its surface. The planets are heated by the radiant energy of the Sun, which gets weaker with the square of the distance. You know how rapidly the

heating effect of a fireplace or an outdoor radiant heater diminishes as you walk away from it; the same effect applies to the Sun. Mercury, the closest planet to the Sun, has a blistering surface temperature that ranges from 280–430 °C on its sunlit side, whereas the surface temperature on Pluto is only about –220 °C, colder than liquid air.

Mathematically, the temperatures decrease approximately in proportion to the square root of the distance from the Sun. Pluto is about 30 AU at its closest to the Sun (or 100 times the distance of Mercury) and about 49 AU at its farthest from the Sun. Thus, Pluto's temperature is less than that of Mercury by the square root of 100, or a factor of 10: from 500 K to 50 K.

In addition to its distance from the Sun, the surface temperature of a planet can be influenced strongly by its atmosphere. Without our atmospheric insulation (the greenhouse effect, which keeps the heat in), the oceans of Earth would be permanently frozen. Conversely, if Mars once had a larger atmosphere in the past, it could have supported a more temperate climate than it has today. Venus is an even more extreme example, where its thick atmosphere of carbon dioxide acts as insulation, reducing the escape of heat built up at the surface, resulting in temperatures greater than those on Mercury. Today, Earth is the only planet where surface temperatures generally lie between the freezing and boiling points of water. As far as we know, Earth is the only planet to support life.

Note:

There's No Place Like Home

In the classic film *The Wizard of Oz*, Dorothy, the heroine, concludes after her many adventures in "alien" environments that "there's no place like home." The same can be said of the other worlds in our solar system. There are many fascinating places, large and small, that we might like to visit, but humans could not survive on any without a great deal of artificial assistance.

A thick carbon dioxide atmosphere keeps the surface temperature on our neighbor Venus at a sizzling 700 K (near 900 °F). Mars, on the other hand, has temperatures generally below freezing, with air (also mostly carbon dioxide) so thin that it resembles that found at an altitude of 30 kilometers

(100,000 feet) in Earth’s atmosphere. And the red planet is so dry that it has not had any rain for billions of years.

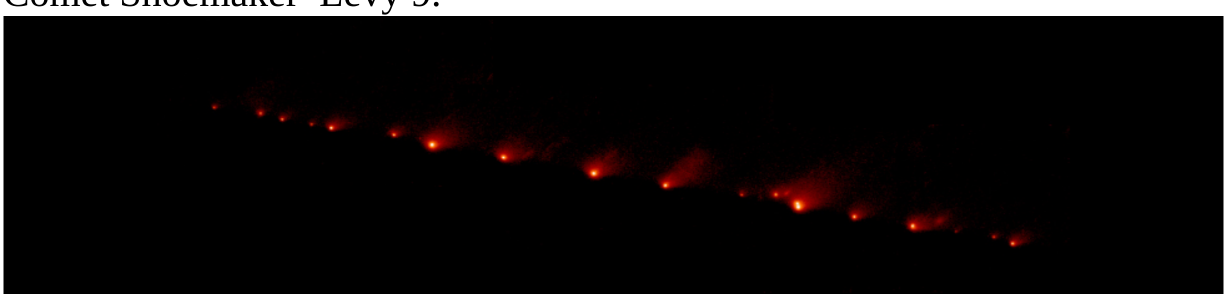
The outer layers of the jovian planets are neither warm enough nor solid enough for human habitation. Any bases we build in the systems of the giant planets may well have to be in space or one of their moons—none of which is particularly hospitable to a luxury hotel with a swimming pool and palm trees. Perhaps we will find warmer havens deep inside the clouds of Jupiter or in the ocean under the frozen ice of its moon Europa.

All of this suggests that we had better take good care of Earth because it is the only site where life as we know it could survive. Recent human activity may be reducing the habitability of our planet by adding pollutants to the atmosphere, especially the potent greenhouse gas carbon dioxide. Human civilization is changing our planet dramatically, and these changes are not necessarily for the better. In a solar system that seems unready to receive us, making Earth less hospitable to life may be a grave mistake.

Geological Activity

The crusts of all of the terrestrial planets, as well as of the larger moons, have been modified over their histories by both internal and external forces. Externally, each has been battered by a slow rain of projectiles from space, leaving their surfaces pockmarked by impact craters of all sizes (see [link]). We have good evidence that this bombardment was far greater in the early history of the solar system, but it certainly continues to this day, even if at a lower rate. The collision of more than 20 large pieces of Comet Shoemaker–Levy 9 with Jupiter in the summer of 1994 (see [link]) is one dramatic example of this process.

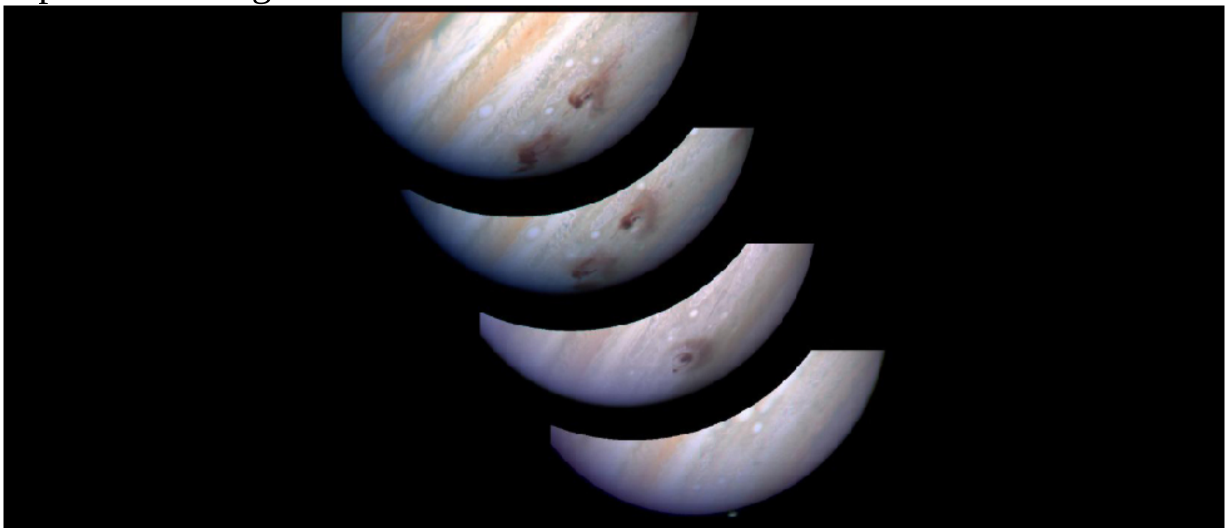
Comet Shoemaker–Levy 9.



In this image of Comet Shoemaker–Levy 9 taken on May 17, 1994, by NASA's Hubble Space Telescope, you can see about 20 icy fragments into which the comet broke. The comet was approximately 660 million kilometers from Earth, heading on a collision course with Jupiter. (credit: modification of work by NASA, ESA, H. Weaver (STScl), E. Smith (STScl))

[link] shows the aftermath of these collisions, when debris clouds larger than Earth could be seen in Jupiter's atmosphere.

Jupiter with Huge Dust Clouds.



The Hubble Space Telescope took this sequence of images of Jupiter in summer 1994, when fragments of Comet Shoemaker–Levy 9 collided with the giant planet. Here we see the site hit by fragment G, from five minutes to five days after impact. Several of the dust clouds generated by the collisions became larger than Earth. (credit: modification of work by H. Hammel, NASA)

During the time all the planets have been subject to such impacts, internal forces on the terrestrial planets have buckled and twisted their crusts, built up mountain ranges, erupted as volcanoes, and generally reshaped the surfaces in what we call geological activity. (The prefix *geo* means "Earth,"

so this is a bit of an “Earth-chauvinist” term, but it is so widely used that we bow to tradition.) Among the terrestrial planets, Earth and Venus have experienced the most geological activity over their histories, although some of the moons in the outer solar system are also surprisingly active. In contrast, our own Moon is a dead world where geological activity ceased billions of years ago.

Geological activity on a planet is the result of a hot interior. The forces of volcanism and mountain building are driven by heat escaping from the interiors of planets. As we will see, each of the planets was heated at the time of its birth, and this primordial heat initially powered extensive volcanic activity, even on our Moon. But, small objects such as the Moon soon cooled off. The larger the planet or moon, the longer it retains its internal heat, and therefore the more we expect to see surface evidence of continuing geological activity. The effect is similar to our own experience with a hot baked potato: the larger the potato, the more slowly it cools. If we want a potato to cool quickly, we cut it into small pieces.

For the most part, the history of volcanic activity on the terrestrial planets conforms to the predictions of this simple theory. The Moon, the smallest of these objects, is a geologically dead world. Although we know less about Mercury, it seems likely that this planet, too, ceased most volcanic activity about the same time the Moon did. Mars represents an intermediate case. It has been much more active than the Moon, but less so than Earth. Earth and Venus, the largest terrestrial planets, still have molten interiors even today, some 4.5 billion years after their birth.

Key Concepts and Summary

The giant planets have dense cores roughly 10 times the mass of Earth, surrounded by layers of hydrogen and helium. The terrestrial planets consist mostly of rocks and metals. They were once molten, which allowed their structures to differentiate (that is, their denser materials sank to the center). The Moon resembles the terrestrial planets in composition, but most of the other moons—which orbit the giant planets—have larger quantities of frozen ice within them. In general, worlds closer to the Sun have higher

surface temperatures. The surfaces of terrestrial planets have been modified by impacts from space and by varying degrees of geological activity.

Glossary

differentiation
: gravitational separation of materials of different density into layers in the interior of a planet or moon

Dating Planetary Surfaces

LEARNING OBJECTIVES

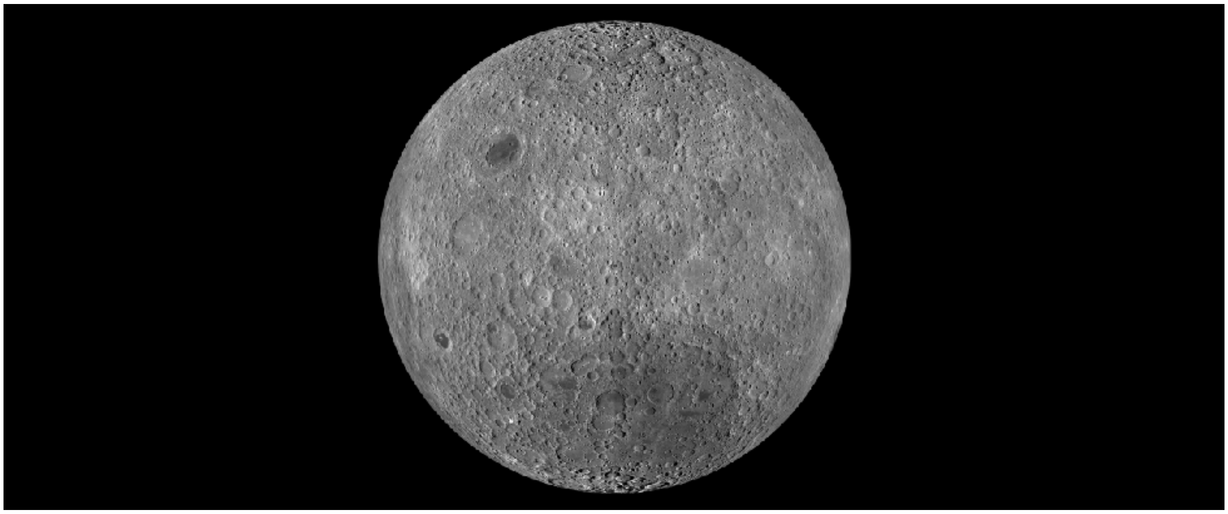
By the end of this section, you will be able to:

- Explain how astronomers can tell whether a planetary surface is geologically young or old
- Describe different methods for dating planets

How do we know the age of the surfaces we see on planets and moons? If a world has a surface (as opposed to being mostly gas and liquid), astronomers have developed some techniques for estimating how long ago that surface solidified. Note that the age of these surfaces is not necessarily the age of the planet as a whole. On geologically active objects (including Earth), vast outpourings of molten rock or the erosive effects of water and ice, which we call planet weathering, have erased evidence of earlier epochs and present us with only a relatively young surface for investigation.

Counting the Craters

One way to estimate the age of a surface is by counting the number of impact craters. This technique works because the rate at which impacts have occurred in the solar system has been roughly constant for several billion years. Thus, in the absence of forces to eliminate craters, the number of craters is simply proportional to the length of time the surface has been exposed. This technique has been applied successfully to many solid planets and moons ([link]).
Our Cratered Moon.



This composite image of the Moon’s surface was made from many smaller images taken between November 2009 and February 2011 by the Lunar Reconnaissance Orbiter (LRO) and shows craters of many different sizes. (credit: modification of work by NASA/GSFC/Arizona State University)

Bear in mind that crater counts can tell us only the time since the surface experienced a major change that could modify or erase preexisting craters. Estimating ages from crater counts is a little like walking along a sidewalk in a snowstorm after the snow has been falling steadily for a day or more. You may notice that in front of one house the snow is deep, while next door the sidewalk may be almost clear. Do you conclude that less snow has fallen in front of Ms. Jones’ house than Mr. Smith’s? More likely, you conclude that Jones has recently swept the walk clean and Smith has not. Similarly, the numbers of craters indicate how long it has been since a planetary surface was last “swept clean” by ongoing lava flows or by molten materials ejected when a large impact happened nearby.

Still, astronomers can use the numbers of craters on different parts of the same world to provide important clues about how regions on that world evolved. On a given planet or moon, the more heavily cratered terrain will generally be older (that is, more time will have elapsed there since something swept the region clean).

Radioactive Rocks

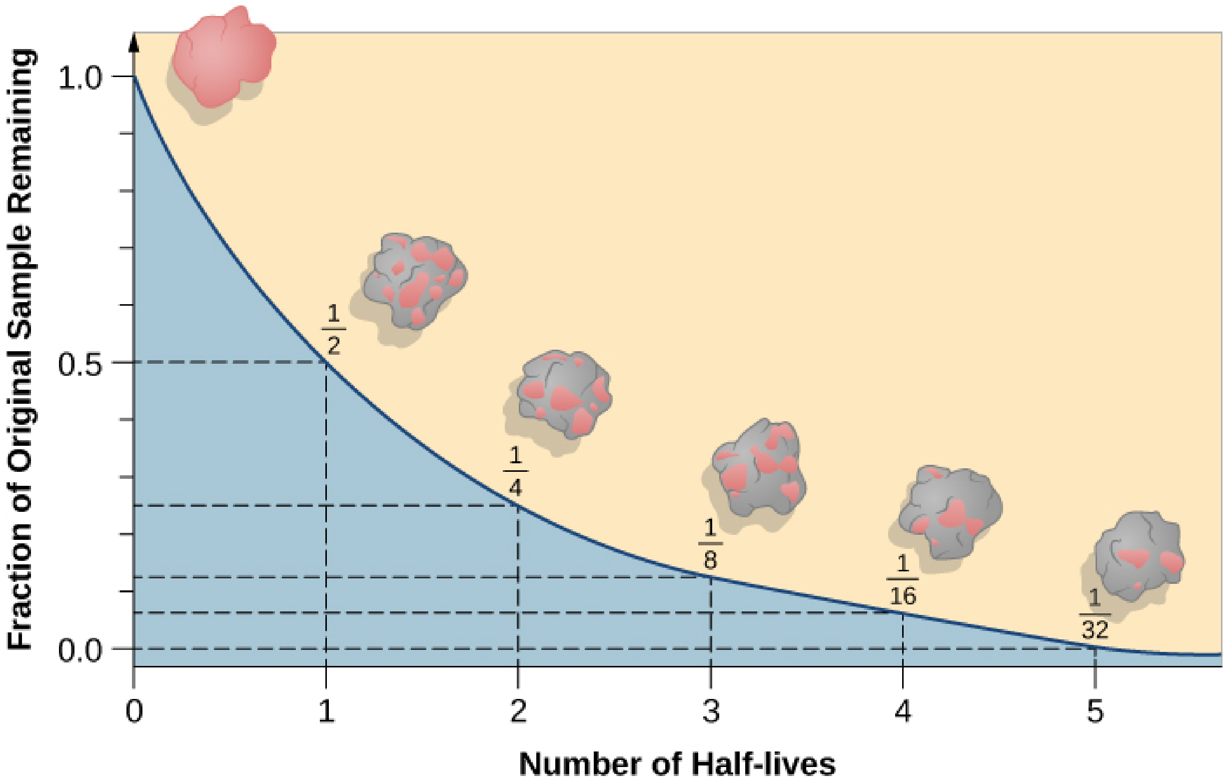
Another way to trace the history of a solid world is to measure the age of individual rocks. After samples were brought back from the Moon by Apollo astronauts, the techniques that had been developed to date rocks on Earth were applied to rock samples from the Moon to establish a geological chronology for the Moon. Furthermore, a few samples of material from the Moon, Mars, and the large asteroid Vesta have fallen to Earth as meteorites and can be examined directly (see the chapter on Cosmic Samples and the Origin of the Solar System).

Scientists measure the age of rocks using the properties of natural **radioactivity**. Around the beginning of the twentieth century, physicists began to understand that some atomic nuclei are not stable but can split apart (decay) spontaneously into smaller nuclei. The process of radioactive decay involves the emission of particles such as electrons, or of radiation in the form of gamma rays (see the chapter on Radiation and Spectra).

For any one radioactive nucleus, it is not possible to predict when the decay process will happen. Such decay is random in nature, like the throw of dice: as gamblers have found all too often, it is impossible to say just when the dice will come up 7 or 11. But, for a very large number of dice tosses, we can calculate the odds that 7 or 11 will come up. Similarly, if we have a very large number of radioactive atoms of one type (say, uranium), there is a specific time period, called its **half-life**, during which the chances are fifty-fifty that decay will occur for any of the nuclei.

A particular nucleus may last a shorter or longer time than its half-life, but in a large sample, almost exactly half of the nuclei will have decayed after a time equal to one half-life. Half of the remaining nuclei will have decayed after two half-lives pass, leaving only one half of a half—or one quarter—of the original sample ([link]).

Radioactive Decay.



This graph shows (in pink) the amount of a radioactive sample that remains after several half-lives have passed. After one half-life, half the sample is left; after two half-lives, one half of the remainder (or one quarter) is left; and after three half-lives, one half of that (or one eighth) is left. Note that, in reality, the decay of radioactive elements in a rock sample would not cause any visible change in the appearance of the rock; the splashes of color are shown here for conceptual purposes only.

If you had 1 gram of pure radioactive nuclei with a half-life of 100 years, then after 100 years you would have
1/2 gram; after 200 years, 1/4 gram; after 300 years, only 1/8 gram; and so forth. However, the material does not disappear. Instead, the radioactive atoms are replaced with their decay products. Sometimes the radioactive atoms are called *parents* and the decay products are called *daughter* elements.

In this way, radioactive elements with half-lives we have determined can provide accurate nuclear clocks. By comparing how much of a radioactive parent element is left in a rock to how much of its daughter products have accumulated, we can learn how long the decay process has been going on and hence how long ago the rock formed. [link] summarizes the decay reactions used most often to date lunar and terrestrial rocks.

Radioactive Decay Reaction Used to Date Rocks[footnote]
The number after each element is its atomic weight, equal to the number of protons plus neutrons in its nucleus. This specifies the *isotope* of the element; different isotopes of the same element differ in the number of neutrons.

Parent	Daughter	Half-Life (billions of years)
Samarium-147	Neodymium-143	106
Rubidium-87	Strontium-87	48.8
Thorium-232	Lead-208	14.0
Uranium-238	Lead-206	4.47
Potassium-40	Argon-40	1.31

Note:
PBS provides an evolution series excerpt that explains how we use radioactive elements to date Earth.

This Science Channel video features Bill Nye the Science Guy showing how scientists have used radioactive dating to determine the age of Earth.

When astronauts first flew to the Moon, one of their most important tasks was to bring back lunar rocks for radioactive age-dating. Until then, astronomers and geologists had no reliable way to measure the age of the lunar surface. Counting craters had let us calculate relative ages (for example, the heavily cratered lunar highlands were older than the dark lava plains), but scientists could not measure the actual age in years. Some thought that the ages were as young as those of Earth's surface, which has been resurfaced by many geological events. For the Moon's surface to be so young would imply active geology on our satellite. Only in 1969, when the first Apollo samples were dated, did we learn that the Moon is an ancient, geologically dead world. Using such dating techniques, we have been able to determine the ages of both Earth and the Moon: each was formed about 4.5 billion years ago (although, as we shall see, Earth probably formed earlier).

We should also note that the decay of radioactive nuclei generally releases energy in the form of heat. Although the energy from a single nucleus is not very large (in human terms), the enormous numbers of radioactive nuclei in a planet or moon (especially early in its existence) can be a significant source of internal energy for that world. Geologists estimate that about half of Earth's current internal heat budget comes from the decay of radioactive isotopes in its interior.

Key Concepts and Summary

The ages of the surfaces of objects in the solar system can be estimated by counting craters: on a given world, a more heavily cratered region will generally be older than one that is less cratered. We can also use samples of rocks with radioactive elements in them to obtain the time since the layer in which the rock formed last solidified. The half-life of a radioactive element is the time it takes for half the sample to decay; we determine how many half-lives have passed by how much of a sample remains the radioactive

element and how much has become the decay product. In this way, we have estimated the age of the Moon and Earth to be roughly 4.5 billion years.

Glossary

half-life
: time required for half of the radioactive atoms in a sample to disintegrate

radioactivity
: process by which certain kinds of atomic nuclei decay naturally, with the spontaneous emission of subatomic particles and gamma rays

Origin of the Solar System

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the characteristics of planets that are used to create formation models of the solar system
- Describe how the characteristics of extrasolar systems help us to model our own solar system
- Explain the importance of collisions in the formation of the solar system

Much of astronomy is motivated by a desire to understand the origin of things: to find at least partial answers to age-old questions of where the universe, the Sun, Earth, and we ourselves came from. Each planet and moon is a fascinating place that may stimulate our imagination as we try to picture what it would be like to visit. Taken together, the members of the solar system preserve patterns that can tell us about the formation of the entire system. As we begin our exploration of the planets, we want to introduce our modern picture of how the solar system formed.

The recent discovery of hundreds of planets in orbit around other stars has shown astronomers that many exoplanetary systems can be quite different from our own solar system. For example, it is common for these systems to include planets intermediate in size between our terrestrial and giant planets. These are often called *superearths*. Some exoplanet systems even have giant planets close to the star, reversing the order we see in our system. In The Birth of Stars and the Discovery of Planets outside the Solar

System, we will look at these exoplanet systems. But for now, let us focus on theories of how our own particular system has formed and evolved.

Looking for Patterns

One way to approach our question of origin is to look for regularities among the planets. We found, for example, that all the planets lie in nearly the same plane and revolve in the same direction around the Sun. The Sun also spins in the same direction about its own axis. Astronomers interpret this pattern as evidence that the Sun and planets formed together from a spinning cloud of gas and dust that we call the **solar nebula** ([link]).

Solar Nebula.

This artist's conception of the solar nebula shows the flattened cloud of gas and dust from which our planetary system formed. Icy and rocky planetesimals (precursors of the planets) can be seen in the

foreground. The bright center is where the Sun is forming. (credit: William K. Hartmann, Planetary Science Institute)

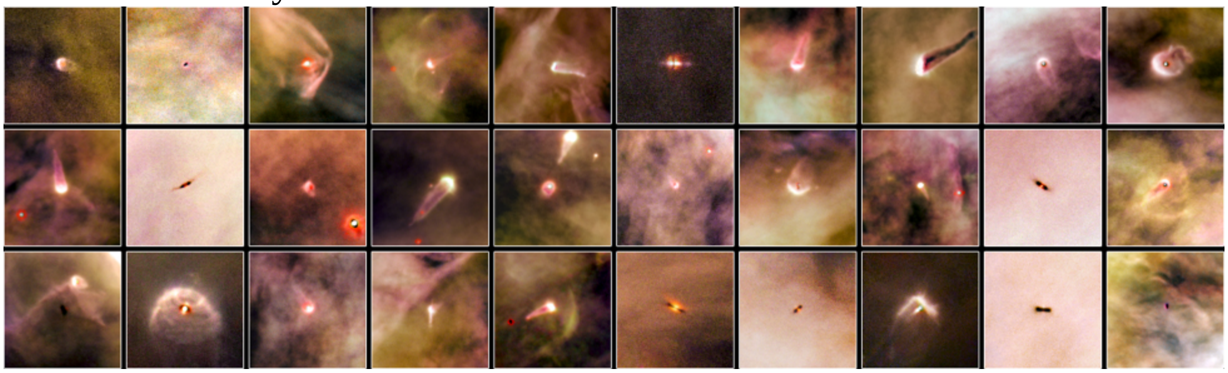
The composition of the planets gives another clue about origins. Spectroscopic analysis allows us to determine which elements are present in the Sun and the planets. The Sun has the same hydrogen-dominated composition as Jupiter and Saturn, and therefore appears to have been formed from the same reservoir of material. In comparison, the terrestrial planets and our Moon are relatively deficient in the light gases and the various ices that form from the common elements oxygen, carbon, and nitrogen. Instead, on Earth and its neighbors, we see mostly the rarer heavy elements such as iron and silicon. This pattern suggests that the processes that led to planet formation in the inner solar system must somehow have excluded much of the lighter materials that are common elsewhere. These lighter materials must have escaped, leaving a residue of heavy stuff.

The reason for this is not hard to guess, bearing in mind the heat of the Sun. The inner planets and most of the asteroids are made of rock and metal, which can survive heat, but they contain very little ice or gas, which evaporate when temperatures are high. (To see what we mean, just compare how long a rock and an ice cube survive when they are placed in the sunlight.) In the outer solar system, where it has always been cooler, the planets and their moons, as well as icy dwarf planets and comets, are composed mostly of ice and gas.

The Evidence from Far Away

A second approach to understanding the origins of the solar system is to look outward for evidence that other systems of planets are forming elsewhere. We cannot look back in time to the formation of our own system, but many stars in space are much younger than the Sun. In these systems, the processes of planet formation might still be accessible to direct observation. We observe that there are many other "solar nebulas" or *circumstellar disks*—flattened, spinning clouds of gas and dust surrounding young stars. These disks resemble our own solar system's initial stages of formation billions of years ago ([link]).

Atlas of Planetary Nurseries.



These Hubble Space Telescope photos show sections of the Orion Nebula, a relatively close-by region where stars are currently forming. Each image shows an embedded circumstellar disk orbiting a very young star. Seen from different angles, some are energized to glow by the light of a nearby star while others are dark and seen in silhouette against the bright glowing gas of the Orion Nebula. Each is a contemporary analog of our own solar nebula—a location where planets are probably being formed today. (credit: modification of work by NASA/ESA, L. Ricci (ESO))

Building Planets

Circumstellar disks are a common occurrence around very young stars, suggesting that disks and stars form together. Astronomers can use theoretical calculations to see how solid bodies might form from the gas and dust in these disks as they cool. These models show that material begins to coalesce first by forming smaller objects, precursors of the planets, which we call **planetesimals**.

Today's fast computers can simulate the way millions of planetesimals, probably no larger than 100 kilometers in diameter, might gather together under their mutual gravity to form the planets we see today. We are beginning to understand that this process was a violent one, with planetesimals crashing into each other and sometimes even disrupting the growing planets themselves. As a consequence of those violent impacts

(and the heat from radioactive elements in them), all the planets were heated until they were liquid and gas, and therefore differentiated, which helps explain their present internal structures.

The process of impacts and collisions in the early solar system was complex and, apparently, often random. The solar nebula model can explain many of the regularities we find in the solar system, but the random collisions of massive collections of planetesimals could be the reason for some exceptions to the “rules” of solar system behavior. For example, why do the planets Uranus and Pluto spin on their sides? Why does Venus spin slowly and in the opposite direction from the other planets? Why does the composition of the Moon resemble Earth in many ways and yet exhibit substantial differences? The answers to such questions probably lie in enormous collisions that took place in the solar system long before life on Earth began.

Today, some 4.5 billion years after its origin, the solar system is—thank goodness—a much less violent place. As we will see, however, some planetesimals have continued to interact and collide, and their fragments move about the solar system as roving “transients” that can make trouble for the established members of the Sun’s family, such as our own Earth. (We discuss this “troublemaking” in Comets and Asteroids: Debris of the Solar System.)

Note:
A great variety of infographics at space.com let you explore what it would be like to live on various worlds in the solar system.

Key Concepts and Summary

Regularities among the planets have led astronomers to hypothesize that the Sun and the planets formed together in a giant, spinning cloud of gas and dust called the solar nebula. Astronomical observations show tantalizingly similar circumstellar disks around other stars. Within the solar nebula,

material first coalesced into planetesimals; many of these gathered together to make the planets and moons. The remainder can still be seen as comets and asteroids. Probably all planetary systems have formed in similar ways, but many exoplanet systems have evolved along quite different paths, as we will see in Cosmic Samples and the Origin of the Solar System.

For Further Exploration

Articles

Davidson, K. “Carl Sagan’s Coming of Age.” *Astronomy.* (November 1999): 40. About the noted popularizer of science and how he developed his interest in astronomy.

Garget, J. “Mysterious Microworlds.” *Astronomy.* (July 2005): 32. A quick tour of a number of the moons in the solar system.

Hartmann, W. “The Great Solar System Revision.” *Astronomy*. (August 1998): 40. How our views have changed over the past 25 years.

Kross, J. “What’s in a Name?” *Sky & Telescope*. (May 1995): 28. How worlds are named.

Rubin, A. “Secrets of Primitive Meteorites.” *Scientific American*. (February 2013): 36. What meteorites can teach us about the environment in which the solar system formed.

Soter, S. “What Is a Planet?” *Scientific American*. (January 2007): 34. The IAU’s new definition of a planet in our solar system, and what happened to Pluto as a result.

Talcott, R. “How the Solar System Came to Be.” *Astronomy*. (November 2012): 24. On the formation period of the Sun and the planets.

Wood, J. “Forging the Planets: The Origin of our Solar System.” *Sky & Telescope*. (January 1999): 36. Good overview.

Websites

Gazetteer of Planetary Nomenclature: http://planetarynames.wr.usgs.gov/. Outlines the rules for naming bodies and features in the solar system.

Planetary Photojournal: http://photojournal.jpl.nasa.gov/index.html. This NASA site features thousands of the best images from planetary exploration, with detailed captions and excellent indexing. You can find images by world, feature name, or mission, and download them in a number of formats. And the images are copyright-free because your tax dollars paid for them.

The following sites present introductory information and pictures about each of the worlds of our solar system:

- NASA/JPL Solar System Exploration pages: http://solarsystem.nasa.gov/index.cfm.
- National Space Science Data Center Lunar and Planetary Science pages: http://nssdc.gsfc.nasa.gov/planetary/.
- Nine [now 8] Planets Solar System Tour: http://www.nineplanets.org/.
- Planetary Society solar system pages: http://www.planetary.org/explore/space-topics/.
- Views of the Solar System by Calvin J. Hamilton: http://www.solarviews.com/eng/homepage.htm.

Videos

Brown Dwarfs and Free Floating Planets: When You Are Just Too Small to Be a Star: https://www.youtube.com/watch?v=zXCDsb4n4KU. A nontechnical talk by Gibor Basri of the University of California at Berkeley, discussing some of the controversies about the meaning of the word "planet" (1:32:52).

In the Land of Enchantment: The Epic Story of the Cassini Mission to Saturn: https://www.youtube.com/watch?v=Vx135n8VFxY. A public lecture by Dr. Carolyn Porco that focuses mainly on the exploration of

Saturn and its moons, but also presents an eloquent explanation of why we explore the solar system (1:37:52).

Origins of the Solar System: http://www.pbs.org/wgbh/nova/space/origins-solar-system.html. A video from PBS that focuses on the evidence from meteorites, narrated by Neil deGrasse Tyson (13:02).

To Scale: The Solar System: https://www.youtube.com/watch?t=84&v=zR3Igc3Rhfg. Constructing a scale model of the solar system in the Nevada desert (7:06).

Collaborative Group Activities

A. Discuss and make a list of the reasons why we humans might want to explore the other worlds in the solar system. Does your group think such missions of exploration are worth the investment? Why?
B. Your instructor will assign each group a world. Your task is to think about what it would be like to be there. (Feel free to look ahead in the book to the relevant chapters.) Discuss where on or around your world we would establish a foothold and what we would need to survive there.
C. In the There's No Place Like Home feature, we discuss briefly how human activity is transforming our planet's overall environment. Can you think of other ways that this is happening?
D. Some scientists criticized Carl Sagan for "wasting his research time" popularizing astronomy. To what extent do you think scientists should spend their time interpreting their field of research for the public? Why or why not? Are there ways that scientists who are not as eloquent or charismatic as Carl Sagan or Neil deGrasse Tyson can still contribute to the public understanding of science?
E. Your group has been named to a special committee by the International Astronomical Union to suggest names of features (such as craters, trenches, and so on) on a newly explored asteroid. Given the restriction that any people after whom features are named must no longer be alive, what names or types of names would you suggest? (Keep in mind that you are not restricted to names of people, by the way.)

F. A member of your group has been kidnapped by a little-known religious cult that worships the planets. They will release him only if your group can tell which of the planets are currently visible in the sky during the evening and morning. You are forbidden from getting your instructor involved. How and where else could you find out the information you need? (Be as specific as you can. If your instructor says it's okay, feel free to answer this question using online or library resources.)
G. In the Carl Sagan: Solar System Advocate feature, you learned that science fiction helped spark and sustain his interest in astronomy. Did any of the members of your group get interested in astronomy as a result of a science fiction story, movie, or TV show? Did any of the stories or films you or your group members saw take place on the planets of our solar system? Can you remember any specific ones that inspired you? If no one in the group is into science fiction, perhaps you can interview some friends or classmates who are and report back to the group.
H. A list of NASA solar system spacecraft missions can be found at http://www.nasa.gov/content/solar-missions-list. Your instructor will assign each group a mission. Look up when the mission was launched and executed, and describe the mission goals, the basic characteristics of the spacecraft (type of instruments, propellant, size, and so on), and what was learned from the mission. If time allows, each group should present its findings to the rest of the class.
I. What would be some of the costs or risks of developing a human colony or base on another planetary body? What technologies would need to be developed? What would people need to give up to live on a different world in our solar system?

Review Questions

Exercise:

Exercise:

Problem:

Venus rotates backward and Uranus and Pluto spin about an axis tipped nearly on its side. Based on what you learned about the motion of small bodies in the solar system and the surfaces of the planets, what might be the cause of these strange rotations?

Exercise:

Problem:

What is the difference between a differentiated body and an undifferentiated body, and how might that influence a body's ability to retain heat for the age of the solar system?

Exercise:

Problem:

What does a planet need in order to retain an atmosphere? How does an atmosphere affect the surface of a planet and the ability of life to exist?

Exercise:

Problem:

Which type of planets have the most moons? Where did these moons likely originate?

Exercise:

Problem: What is the difference between a meteor and a meteorite?

Exercise:

Problem:

Explain our ideas about why the terrestrial planets are rocky and have less gas than the giant planets.

Exercise:

Problem: Do all planetary systems look the same as our own?

Exercise:

Problem:

What is comparative planetology and why is it useful to astronomers?

Exercise:

Problem:

What changed in our understanding of the Moon and Moon-Earth system as a result of humans landing on the Moon’s surface?

Exercise:

Problem:

If Earth was to be hit by an extraterrestrial object, where in the solar system could it come from and how would we know its source region?

Exercise:

Problem:

List some reasons that the study of the planets has progressed more in the past few decades than any other branch of astronomy.

Exercise:

Problem:

Imagine you are a travel agent in the next century. An eccentric billionaire asks you to arrange a “Guinness Book of Solar System Records” kind of tour. Where would you direct him to find the following (use this chapter and Appendix F and Appendix G):

A. the least-dense planet
B. the densest planet
C. the largest moon in the solar system

D. excluding the jovian planets, the planet where you would weigh the most on its surface (Hint: Weight is directly proportional to surface gravity.)
E. the smallest planet
F. the planet that takes the longest time to rotate
G. the planet that takes the shortest time to rotate
H. the planet with a diameter closest to Earth’s
I. the moon with the thickest atmosphere
J. the densest moon
K. the most massive moon

Exercise:

Problem:

What characteristics do the worlds in our solar system have in common that lead astronomers to believe that they all formed from the same “mother cloud” (solar nebula)?

Exercise:

Problem:

How do terrestrial and giant planets differ? List as many ways as you can think of.

Exercise:

Problem:

Why are there so many craters on the Moon and so few on Earth?

Exercise:

Problem: How do asteroids and comets differ?

Exercise:

Problem:

How and why is Earth’s Moon different from the larger moons of the giant planets?

Exercise:

Problem:

Where would you look for some “original” planetesimals left over from the formation of our solar system?

Exercise:

Problem:

Describe how we use radioactive elements and their decay products to find the age of a rock sample. Is this necessarily the age of the entire world from which the sample comes? Explain.

Exercise:

Problem:

What was the solar nebula like? Why did the Sun form at its center?

Thought Questions

Exercise:

Problem:

What can we learn about the formation of our solar system by studying other stars? Explain.

Exercise:

Problem:

Earlier in this chapter, we modeled the solar system with Earth at a distance of about one city block from the Sun. If you were to make a model of the distances in the solar system to match your height, with the Sun at the top of your head and Pluto at your feet, which planet would be near your waist? How far down would the zone of the terrestrial planets reach?

Exercise:

Problem:

Seasons are a result of the inclination of a planet's axial tilt being inclined from the normal of the planet's orbital plane. For example, Earth has an axis tilt of 23.4° (Appendix F). Using information about just the inclination alone, which planets might you expect to have seasonal cycles similar to Earth, although different in duration because orbital periods around the Sun are different?

Exercise:

Problem:

Again using Appendix F, which planet(s) might you expect not to have significant seasonal activity? Why?

Exercise:

Problem:

Again using Appendix F, which planets might you expect to have extreme seasons? Why?

Exercise:

Problem:

Using some of the astronomical resources in your college library or the Internet, find five names of features on each of three other worlds that are named after real people. In a sentence or two, describe each of these people and what contributions they made to the progress of science or human thought.

Exercise:

Problem:

Explain why the planet Venus is differentiated, but asteroid Fraknoi, a very boring and small member of the asteroid belt, is not.

Exercise:

Problem:

Would you expect as many impact craters per unit area on the surface of Venus as on the surface of Mars? Why or why not?

Exercise:

Problem:

Interview a sample of 20 people who are not taking an astronomy class and ask them if they can name a living astronomer. What percentage of those interviewed were able to name one? Typically, the two living astronomers the public knows these days are Stephen Hawking and Neil deGrasse Tyson. Why are they better known than most astronomers? How would your result have differed if you had asked the same people to name a movie star or a professional basketball player?

Exercise:

Problem:

Using Appendix G, complete the following table that describes the characteristics of the Galilean moons of Jupiter, starting from Jupiter and moving outward in distance.

Moon	Semimajor Axis (km^3)	Diameter	Density (g/cm^3)
Io			
Europa			
Ganymede			

Moon	Semimajor Axis (km^3)	Diameter	Density (g/cm^3)
Callisto			

Table A

This system has often been described as a mini solar system. Why might this be so? If Jupiter were to represent the Sun and the Galilean moons represented planets, which moons could be considered more terrestrial in nature and which ones more like gas/ice giants? Why? (Hint: Use the values in your table to help explain your categorization.)

Figuring for Yourself

Exercise:

Problem:

Calculate the density of Jupiter. Show your work. Is it more or less dense than Earth? Why?

Exercise:

Problem:

Calculate the density of Saturn. Show your work. How does it compare with the density of water? Explain how this can be.

Exercise:

Problem:

What is the density of Jupiter’s moon Europa (see Appendix G for data on moons)? Show your work.

Exercise:

Problem:

Look at Appendix F and Appendix G and indicate the moon with a diameter that is the largest fraction of the diameter of the planet or dwarf planet it orbits.

Exercise:

Problem:

Barnard’s Star, the second closest star to us, is about 56 trillion (5.6×10^{12}) km away. Calculate how far it would be using the scale model of the solar system given in Overview of Our Planetary System.

Exercise:

Problem:

A radioactive nucleus has a half-life of 5×10^8 years. Assuming that a sample of rock (say, in an asteroid) solidified right after the solar system formed, approximately what fraction of the radioactive element should be left in the rock today?

Glossary

planetesimals
: objects, from tens to hundreds of kilometers in diameter, that formed in the solar nebula as an intermediate step between tiny grains and the larger planetary objects we see today; the comets and some asteroids may be leftover planetesimals

solar nebula
: the cloud of gas and dust from which the solar system formed

Thinking Ahead
class="introduction"
Active Geology.

This image, taken from the International Space Station in 2006, shows a plume of ash coming from the Cleveland Volcano in the Aleutian Islands. Although the plume was only visible for around two hours, such events are a testament to the dynamic nature of Earth’s crust. (credit: modification of work by NASA)

Airless worlds in our solar system seem peppered with craters large and small. Earth, on the other hand, has few craters, but a thick atmosphere and much surface activity. Although impacts occurred on Earth at the same rate, craters have since been erased by forces in the planet’s crust and atmosphere. What can the comparison between the obvious persistent cratering on so many other worlds, and the different appearance of Earth, tell us about the history of our planet?

As our first step in exploring the solar system in more detail, we turn to the most familiar planet, our own Earth. The first humans to see Earth as a blue sphere floating in the blackness of space were the astronauts who made the first voyage around the Moon in 1968. For many people, the historic images showing our world as a small, distant globe represent a pivotal moment in human history, when it became difficult for educated human beings to view our world without a global perspective. In this chapter, we examine the composition and structure of our planet with its envelope of ocean and atmosphere. We ask how our terrestrial environment came to be the way it is today, and how it compares with other planets.

The Global Perspective

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the components of Earth’s interior and explain how scientists determined its structure
- Specify the origin, size, and extent of Earth’s magnetic field

Earth is a medium-size planet with a diameter of approximately 12,760 kilometers ([link]). As one of the inner or terrestrial planets, it is composed primarily of heavy elements such as iron, silicon, and oxygen—very different from the composition of the Sun and stars, which are dominated by the light elements hydrogen and helium. Earth’s orbit is nearly circular, and Earth is warm enough to support liquid water on its surface. It is the only planet in our solar system that is neither too hot nor too cold, but “just right” for the development of life as we know it. Some of the basic properties of Earth are summarized in [link].

Blue Marble.

This image of Earth from space, taken by the Apollo 17 astronauts, is known as the “Blue Marble.” This is one of the rare images of a full Earth taken during the Apollo program; most images show only part of Earth’s disk in sunlight. (credit: modification of work by NASA)

Some Properties of Earth	
Property	**Measurement**
Semimajor axis	1.00 AU
Period	1.00 year
Mass	5.98×10^{24} kg
Diameter	12,756 km

Some Properties of Earth	
Property	**Measurement**
Radius	6378 km
Escape velocity	11.2 km/s
Rotational period	23 h 56 m 4 s
Surface area	5.1×10^8 km^2
Density	5.514 g/cm^3
Atmospheric pressure	1.00 bar

Earth's Interior

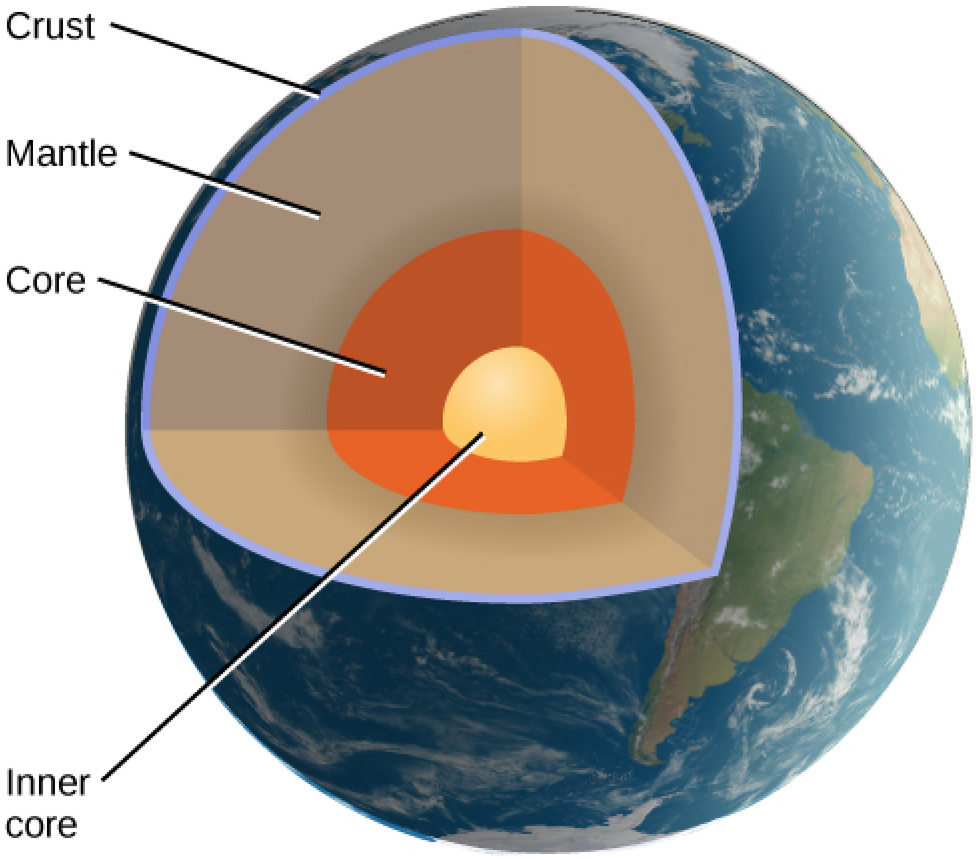
The interior of a planet—even our own Earth—is difficult to study, and its composition and structure must be determined indirectly. Our only direct experience is with the outermost skin of Earth's crust, a layer no more than a few kilometers deep. It is important to remember that, in many ways, we know less about our own planet 5 kilometers beneath our feet than we do about the surfaces of Venus and Mars.

Earth is composed largely of metal and silicate rock (see the Composition and Structure of Planets section). Most of this material is in a solid state, but some of it is hot enough to be molten. The structure of material in Earth's interior has been probed in considerable detail by measuring the transmission of **seismic waves** through Earth. These are waves that spread through the interior of Earth from earthquakes or explosion sites.

Seismic waves travel through a planet rather like sound waves through a struck bell. Just as the sound frequencies vary depending on the material the bell is made of and how it is constructed, so a planet's response depends on its composition and structure. By monitoring the seismic waves in different

locations, scientists can learn about the layers through which the waves have traveled. Some of these vibrations travel along the surface; others pass directly through the interior. Seismic studies have shown that Earth's interior consists of several distinct layers with different compositions, illustrated in [link]. As waves travel through different materials in Earth's interior, the waves—just like light waves in telescope lenses—bend (or refract) so that some seismic stations on Earth receive the waves and others are in "shadows." Detecting the waves in a network of seismographs helps scientists construct a model of Earth's interior, showing liquid and solid layers. This type of seismic imaging is not unlike that used in ultrasound, a type of imaging used to see inside the body.

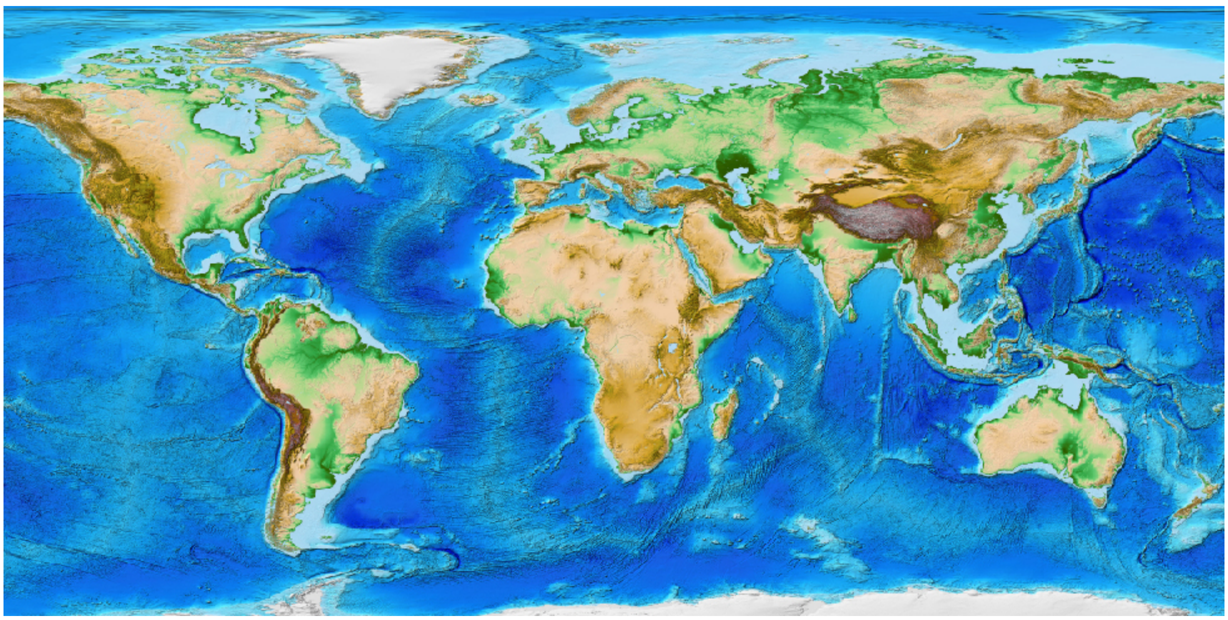
Interior Structure of Earth.



The crust, mantle, and inner and outer cores (solid and liquid, respectively) as shown as revealed by seismic studies.

The top layer is the **crust**, the part of Earth we know best ([link]). Oceanic crust covers 55% of Earth’s surface and lies mostly submerged under the oceans. It is typically about 6 kilometers thick and is composed of volcanic rocks called **basalt**. Produced by the cooling of volcanic lava, basalts are made primarily of the elements silicon, oxygen, iron, aluminum, and magnesium. The continental crust covers 45% of the surface, some of which is also beneath the oceans. The continental crust is 20 to 70 kilometers thick and is composed predominantly of a different volcanic class of silicates (rocks made of silicon and oxygen) called **granite**. These crustal rocks, both oceanic and continental, typically have densities of about 3 g/cm^3. (For comparison, the density of water is 1 g/cm^3.) The crust is the easiest layer for geologists to study, but it makes up only about 0.3% of the total mass of Earth.

Earth’s Crust.



This computer-generated image shows the surface of Earth’s crust as determined from satellite images and ocean floor radar mapping. Oceans and lakes are shown in blue, with darker areas representing depth. Dry land is shown in shades of green and brown, and the Greenland and Antarctic ice sheets are depicted in shades of white. (credit: modification of work by C. Amante, B. W. Eakins, National Geophysical Data Center, NOAA)

The largest part of the solid Earth, called the **mantle**, stretches from the base of the crust downward to a depth of 2900 kilometers. The mantle is more or less solid, but at the temperatures and pressures found there, mantle rock can deform and flow slowly. The density in the mantle increases downward from about 3.5 g/cm^3 to more than 5 g/cm^3 as a result of the compression produced by the weight of overlying material. Samples of upper mantle material are occasionally ejected from volcanoes, permitting a detailed analysis of its chemistry.

Beginning at a depth of 2900 kilometers, we encounter the dense metallic **core** of Earth. With a diameter of 7000 kilometers, our core is substantially larger than the entire planet Mercury. The outer core is liquid, but the innermost part of the core (about 2400 kilometers in diameter) is probably solid. In addition to iron, the core probably also contains substantial quantities of nickel and sulfur, all compressed to a very high density.

The separation of Earth into layers of different densities is an example of *differentiation*, the process of sorting the major components of a planet by density. The fact that Earth is differentiated suggests that it was once warm enough for its interior to melt, permitting the heavier metals to sink to the center and form the dense core. Evidence for differentiation comes from comparing the planet’s bulk density (5.5 g/cm^3) with the surface materials (3 g/cm^3) to suggest that denser material must be buried in the core.

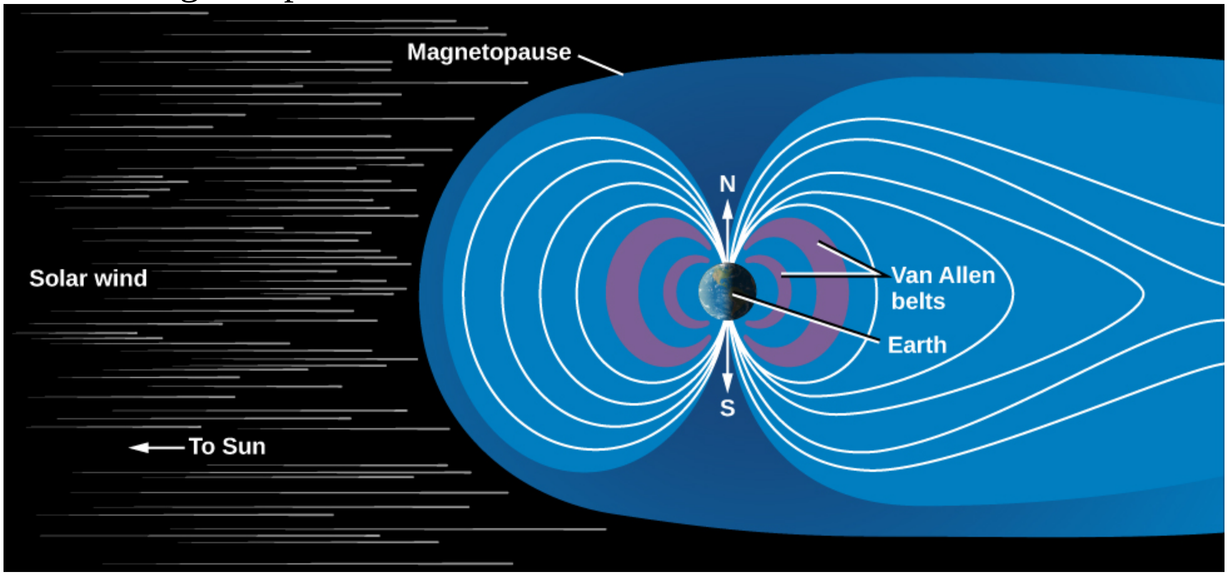
Magnetic Field and Magnetosphere

We can find additional clues about Earth’s interior from its magnetic field. Our planet behaves in some ways as if a giant bar magnet were inside it, aligned approximately with the rotational poles of Earth. This magnetic field is generated by moving material in Earth’s liquid metallic core. As the liquid metal inside Earth circulates, it sets up a circulating electric current. When many charged particles are moving together like that—in the laboratory or on the scale of an entire planet—they produce a magnetic field.

Earth’s magnetic field extends into surrounding space. When a charged particle encounters a magnetic field in space, it becomes trapped in the

magnetic zone. Above Earth’s atmosphere, our field is able to trap small quantities of electrons and other atomic particles. This region, called the **magnetosphere**, is defined as the zone within which Earth’s magnetic field dominates over the weak interplanetary magnetic field that extends outward from the Sun ([link]).

Earth’s Magnetosphere.



A cross-sectional view of our magnetosphere (or zone of magnetic influence), as revealed by numerous spacecraft missions. Note how the wind of charged particles from the Sun “blows” the magnetic field outward like a wind sock.

Where do the charged particles trapped in our magnetosphere come from? They flow outward from the hot surface of the Sun; this is called the *solar wind*. It not only provides particles for Earth’s magnetic field to trap, it also stretches our field in the direction pointing away from the Sun. Typically, Earth’s magnetosphere extends about 60,000 kilometers, or 10 Earth radii, in the direction of the Sun. But, in the direction away from the Sun, the magnetic field can reach as far as the orbit of the Moon, and sometimes farther.

The magnetosphere was discovered in 1958 by instruments on the first US Earth satellite, *Explorer 1*, which recorded the ions (charged particles)

trapped in its inner part. The regions of high-energy ions in the magnetosphere are often called the *Van Allen belts* in recognition of the University of Iowa professor who built the scientific instrumentation for *Explorer 1*. Since 1958, hundreds of spacecraft have explored various regions of the magnetosphere. You can read more about its interaction with the Sun in a later chapter.

Key Concepts and Summary

Earth is the prototype terrestrial planet. Its interior composition and structure are probed using seismic waves. Such studies reveal that Earth has a metal core and a silicate mantle. The outer layer, or crust, consists primarily of oceanic basalt and continental granite. A global magnetic field, generated in the core, produces Earth’s magnetosphere, which can trap charged atomic particles.

Glossary

basalt
: igneous rock produced by the cooling of lava; makes up most of Earth’s oceanic crust and is found on other planets that have experienced extensive volcanic activity

core
: the central part of the planet; consists of higher density material

crust
: the outer layer of a terrestrial planet

granite
: a type of igneous silicate rock that makes up most of Earth’s continental crust

magnetosphere
: the region around a planet in which its intrinsic magnetic field dominates the interplanetary field carried by the solar wind; hence, the

region within which charged particles can be trapped by the planetary magnetic field

mantle
: the largest part of Earth’s interior; lies between the crust and the core

seismic wave
: a vibration that travels through the interior of Earth or any other object; on Earth, these are generally caused by earthquakes

Earth’s Crust

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Denote the primary types of rock that constitute Earth’s crust
- Explain the theory of plate tectonics
- Describe the difference between rift and subduction zones
- Describe the relationship between fault zones and mountain building
- Explain the various types of volcanic activity occurring on Earth

Let us now examine our planet’s outer layers in more detail. Earth’s crust is a dynamic place. Volcanic eruptions, erosion, and large-scale movements of the continents rework the surface of our planet constantly. Geologically, ours is the most active planet. Many of the geological processes described in this section have taken place on other planets as well, but usually in their distant pasts. Some of the moons of the giant planets also have impressive activity levels. For example, Jupiter’s moon Io has a remarkable number of active volcanoes.

Composition of the Crust

Earth’s crust is largely made up of oceanic basalt and continental granite. These are both **igneous rock**, the term used for any rock that has cooled from a molten state. All volcanically produced rock is igneous ([link]).
Formation of Igneous Rock as Liquid Lava Cools and Freezes.

This is a lava flow from a basaltic eruption. Basaltic lava flows quickly and can move easily over distances of more than 20 kilometers. (credit: USGS)

Two other kinds of rock are familiar to us on Earth, although it turns out that neither is common on other planets. **Sedimentary rocks** are made of fragments of igneous rock or the shells of living organisms deposited by wind or water and cemented together without melting. On Earth, these rocks include the common sandstones, shales, and limestones. **Metamorphic rocks** are produced when high temperature or pressure alters igneous or sedimentary rock physically or chemically (the word *metamorphic* means "changed in form"). Metamorphic rocks are produced on Earth because geological activity carries surface rocks down to considerable depths and then brings them back up to the surface. Without such activity, these changed rocks would not exist at the surface.

There is a fourth very important category of rock that can tell us much about the early history of the planetary system: **primitive rock**, which has largely escaped chemical modification by heating. Primitive rock represents the original material out of which the planetary system was made. No primitive material is left on Earth because the entire planet was heated early

in its history. To find primitive rock, we must look to smaller objects such as comets, asteroids, and small planetary moons. We can sometimes see primitive rock in samples that fall to Earth from these smaller objects.

A block of quartzite on Earth is composed of materials that have gone through all four of these states. Beginning as primitive material before Earth was born, it was heated in the early Earth to form igneous rock, transformed chemically and redeposited (perhaps many times) to form sedimentary rock, and finally changed several kilometers below Earth's surface into the hard, white metamorphic stone we see today.

Plate Tectonics

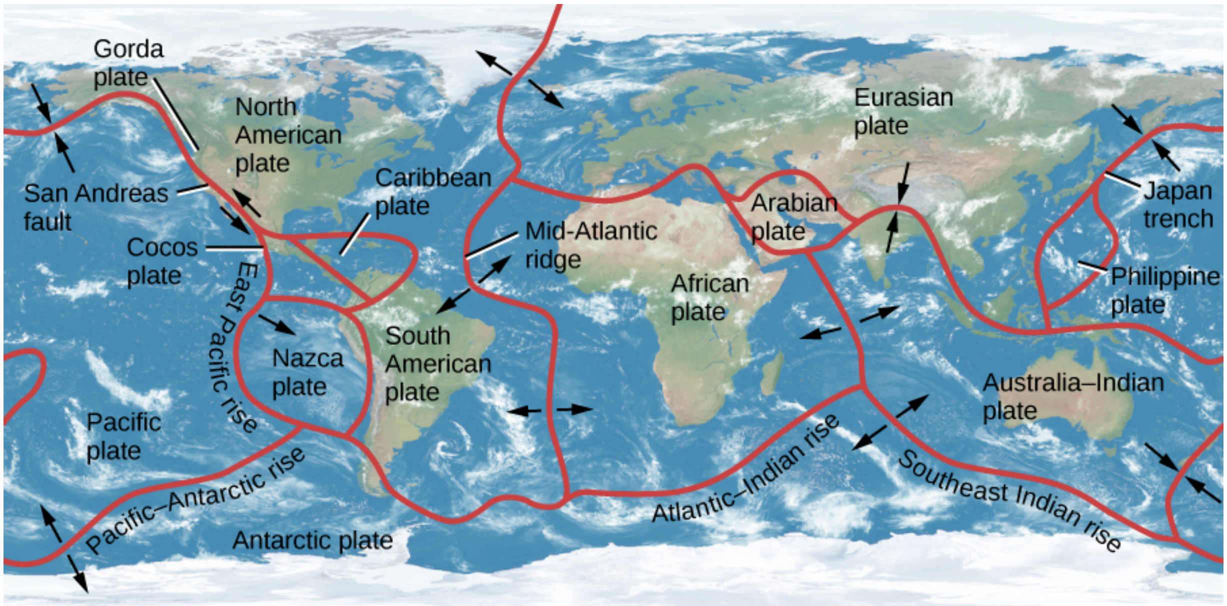
Geology is the study of Earth's crust and the processes that have shaped its surface throughout history. (Although *geo-* means "related to Earth," astronomers and planetary scientists also talk about the geology of other planets.) Heat escaping from the interior provides energy for the formation of our planet's mountains, valleys, volcanoes, and even the continents and ocean basins themselves. But not until the middle of the twentieth century did geologists succeed in understanding just how these landforms are created.

Plate tectonics is a theory that explains how slow motions within the mantle of Earth move large segments of the crust, resulting in a gradual "drifting" of the continents as well as the formation of mountains and other large-scale geological features. Plate tectonics is a concept as basic to geology as evolution by natural selection is to biology or gravity is to understanding the orbits of planets. Looking at it from a different perspective, plate tectonics is a mechanism for Earth to transport heat efficiently from the interior, where it has accumulated, out to space. It is a cooling system for the planet. All planets develop a heat transfer process as they evolve; mechanisms may differ from that on Earth as a result of chemical makeup and other constraints.

Earth's crust and upper mantle (to a depth of about 60 kilometers) are divided into about a dozen tectonic plates that fit together like the pieces of a jigsaw puzzle ([link]). In some places, such as the Atlantic Ocean, the

plates are moving apart; in others, such as off the western coast of South America, they are being forced together. The power to move the plates is provided by slow **convection** of the mantle, a process by which heat escapes from the interior through the upward flow of warmer material and the slow sinking of cooler material. (Convection, in which energy is transported from a warm region, such as the interior of Earth, to a cooler region, such as the upper mantle, is a process we encounter often in astronomy—in stars as well as planets. It is also important in boiling water for coffee while studying for astronomy exams.)

Earth’s Continental Plates.



This map shows the major plates into which the crust of Earth is divided. Arrows indicate the motion of the plates at average speeds of 4 to 5 centimeters per year, similar to the rate at which your hair grows.

Note:
The US Geological Survey provides a map of recent earthquakes and shows the boundaries of the tectonic plates and where earthquakes occur in

relation to these boundaries. You can look close-up at the United States or zoom out for a global view.

As the plates slowly move, they bump into each other and cause dramatic changes in Earth's crust over time. Four basic kinds of interactions between crustal plates are possible at their boundaries: (1) they can pull apart, (2) one plate can burrow under another, (3) they can slide alongside each other, or (4) they can jam together. Each of these activities is important in determining the geology of Earth.

Note:
Alfred Wegener: Catching the Drift of Plate Tectonics
When studying maps or globes of Earth, many students notice that the coast of North and South America, with only minor adjustments, could fit pretty well against the coast of Europe and Africa. It seems as if these great landmasses could once have been together and then were somehow torn apart. The same idea had occurred to others (including Francis Bacon as early as 1620), but not until the twentieth century could such a proposal be more than speculation. The scientist who made the case for continental drift in 1920 was a German meteorologist and astronomer named Alfred Wegener ([link]).
Alfred Wegener (1880–1930).

Wegener proposed a scientific theory for the slow shifting of the continents.

Born in Berlin in 1880, Wegener was, from an early age, fascinated by Greenland, the world’s largest island, which he dreamed of exploring. He studied at the universities in Heidelberg, Innsbruck, and Berlin, receiving a doctorate in astronomy by reexamining thirteenth-century astronomical tables. But, his interests turned more and more toward Earth, particularly its weather. He carried out experiments using kites and balloons, becoming so accomplished that he and his brother set a world record in 1906 by flying for 52 hours in a balloon.

Wegener first conceived of continental drift in 1910 while examining a world map in an atlas, but it took 2 years for him to assemble sufficient data to propose the idea in public. He published the results in book form in 1915. Wegener's evidence went far beyond the congruence in the shapes of the continents. He proposed that the similarities between fossils found only in South America and Africa indicated that these two continents were joined at one time. He also showed that resemblances among living animal species on different continents could best be explained by assuming that the continents were once connected in a supercontinent he called *Pangaea* (from Greek elements meaning "all land").

Wegener's suggestion was met with a hostile reaction from most scientists. Although he had marshaled an impressive list of arguments for his hypothesis, he was missing a *mechanism*. No one could explain *how* solid continents could drift over thousands of miles. A few scientists were sufficiently impressed by Wegener's work to continue searching for additional evidence, but many found the notion of moving continents too revolutionary to take seriously. Developing an understanding of the mechanism (plate tectonics) would take decades of further progress in geology, oceanography, and geophysics.

Wegener was disappointed in the reception of his suggestion, but he continued his research and, in 1924, he was appointed to a special meteorology and geophysics professorship created especially for him at the University of Graz (where he was, however, ostracized by most of the geology faculty). Four years later, on his fourth expedition to his beloved Greenland, he celebrated his fiftieth birthday with colleagues and then set off on foot toward a different camp on the island. He never made it; he was found a few days later, dead of an apparent heart attack.

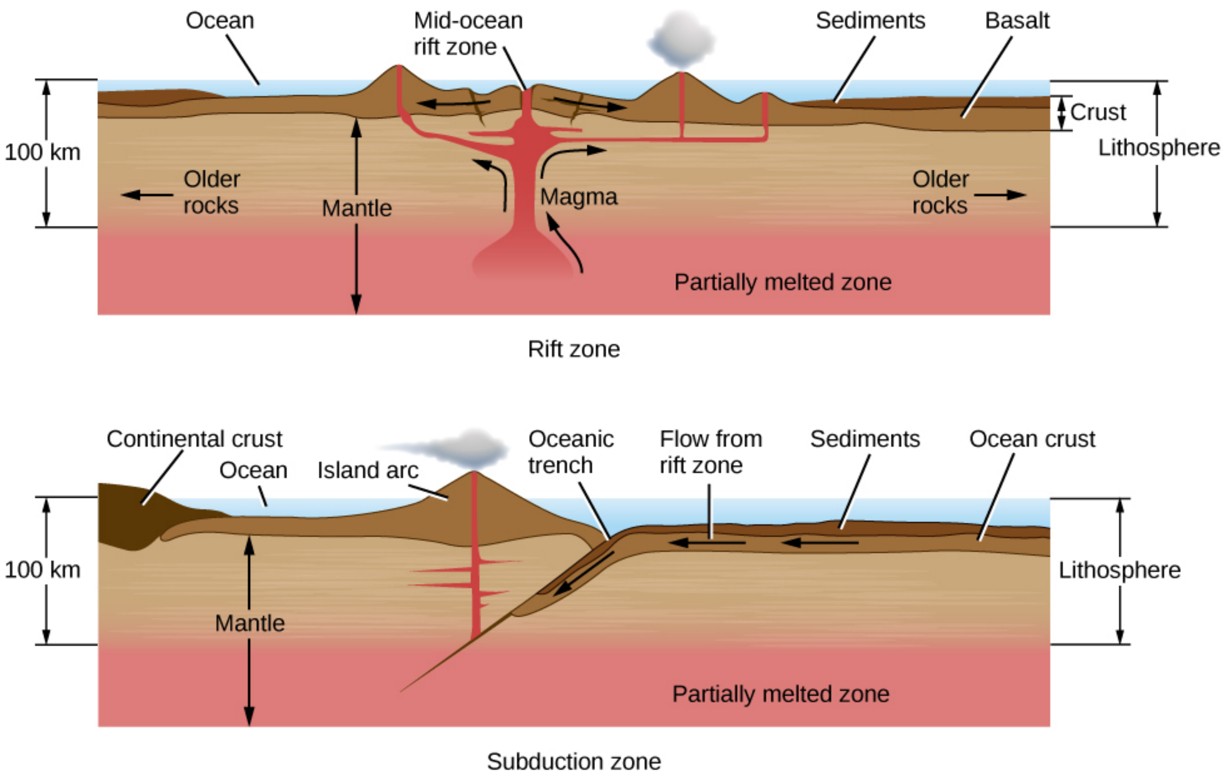
Critics of science often point to the resistance to the continental drift hypothesis as an example of the flawed way that scientists regard new ideas. (Many people who have advanced crackpot theories have claimed that they are being ridiculed unjustly, just as Wegener was.) But we think there is a more positive light in which to view the story of Wegener's suggestion. Scientists in his day maintained a skeptical attitude because they needed more evidence and a clear mechanism that would fit what they understood about nature. Once the evidence and the mechanism were clear, Wegener's hypothesis quickly became the centerpiece of our view of a dynamic Earth.

Note:
See how the [drift of the continents](#) has changed the appearance of our planet’s crust.

Rift and Subduction Zones

Plates pull apart from each other along **rift zones**, such as the Mid-Atlantic ridge, driven by upwelling currents in the mantle ([link]). A few rift zones are found on land. The best known is the central African rift—an area where the African continent is slowly breaking apart. Most rift zones, however, are in the oceans. Molten rock rises from below to fill the space between the receding plates; this rock is basaltic lava, the kind of igneous rock that forms most of the ocean basins.

Rift Zone and Subduction Zone.



Rift and subduction zones are the regions (mostly beneath the oceans) where new crust is formed and old crust is destroyed as part of the cycle of plate tectonics.

From a knowledge of how the seafloor is spreading, we can calculate the average age of the oceanic crust. About 60,000 kilometers of active rifts have been identified, with average separation rates of about 5 centimeters per year. The new area added to Earth each year is about 2 square kilometers, enough to renew the entire oceanic crust in a little more than 100 million years. This is a very short interval in geological time—less than 3% of the age of Earth. The present ocean basins thus turn out to be among the youngest features on our planet.

As new crust is added to Earth, the old crust must go somewhere. When two plates come together, one plate is often forced beneath another in what is called a **subduction** zone ([link]). In general, the thick continental masses cannot be subducted, but the thinner oceanic plates can be rather readily thrust down into the upper mantle. Often a subduction zone is marked by an ocean trench; a fine example of this type of feature is the deep Japan trench along the coast of Asia. The subducted plate is forced down into regions of high pressure and temperature, eventually melting several hundred kilometers below the surface. Its material is recycled into a downward-flowing convection current, ultimately balancing the flow of material that rises along rift zones. The amount of crust destroyed at subduction zones is approximately equal to the amount formed at rift zones.

All along the subduction zone, earthquakes and volcanoes mark the death throes of the plate. Some of the most destructive earthquakes in history have taken place along subduction zones, including the 1923 Yokohama earthquake and fire that killed 100,000 people, the 2004 Sumatra earthquake and tsunami that killed more than 200,000 people, and the 2011 Tohoku earthquake that resulted in the meltdown of three nuclear power reactors in Japan.

Fault Zones and Mountain Building

Along much of their length, the crustal plates slide parallel to each other. These plate boundaries are marked by cracks or **faults**. Along active fault

zones, the motion of one plate with respect to the other is several centimeters per year, about the same as the spreading rates along rifts.

One of the most famous faults is the San Andreas Fault in California, which lies at the boundary between the Pacific plate and the North American plate ([link]). This fault runs from the Gulf of California to the Pacific Ocean northwest of San Francisco. The Pacific plate, to the west, is moving northward, carrying Los Angeles, San Diego, and parts of the southern California coast with it. In several million years, Los Angeles may be an island off the coast of San Francisco.

San Andreas Fault.

We see part of a very active region in California where one crustal plate is sliding sideways with respect to the other. The fault is marked by the valley running up the right side of the photo. Major slippages along this fault can produce extremely destructive earthquakes. (credit: John Wiley)

Unfortunately for us, the motion along fault zones does not take place smoothly. The creeping motion of the plates against each other builds up stresses in the crust that are released in sudden, violent slippages that generate earthquakes. Because the average motion of the plates is constant, the longer the interval between earthquakes, the greater the stress and the more energy released when the surface finally moves.

For example, the part of the San Andreas Fault near the central California town of Parkfield has slipped every 25 years or so during the past century, moving an average of about 1 meter each time. In contrast, the average interval between major earthquakes in the Los Angeles region is about 150 years, and the average motion is about 7 meters. The last time the San Andreas fault slipped in this area was in 1857; tension has been building ever since, and sometime soon it is bound to be released. Sensitive instruments placed within the Los Angeles basin show that the basin is distorting and contracting in size as these tremendous pressures build up beneath the surface.

Example:

Fault Zones and Plate Motion

After scientists mapped the boundaries between tectonic plates in Earth's crust and measured the annual rate at which the plates move (which is about 5 cm/year), we could estimate quite a lot about the rate at which the geology of Earth is changing. As an example, let's suppose that the next slippage along the San Andreas Fault in southern California takes place in the year 2017 and that it completely relieves the accumulated strain in this region. How much slippage is required for this to occur?

Solution

The speed of motion of the Pacific plate relative to the North American plate is 5 cm/y. That's 500 cm (or 5 m) per century. The last southern California earthquake was in 1857. The time from 1857 to 2017 is 160 y, or 1.6 centuries, so the slippage to relieve the strain completely

would be
5 m/century × 1.6 centuries = 8.0 m.

Check Your Learning

If the next major southern California earthquake occurs in 2047 and only relieves one-half of the accumulated strain, how much slippage will occur?

Note:

Answer:

The difference in time from 1857 to 2047 is 190 y, or 1.9 centuries. Because only half the strain is released, this is equivalent to half the annual rate of motion. The total slippage comes to
0.5 × 5 m/century × 1.9 centuries = 4.75 m.

When two continental masses are moving on a collision course, they push against each other under great pressure. Earth buckles and folds, dragging some rock deep below the surface and raising other folds to heights of many kilometers. This is the way many, but not all, of the mountain ranges on Earth were formed. The Alps, for example, are a result of the African plate bumping into the Eurasian plate. As we will see, however, quite different processes produced the mountains on other planets.

Once a mountain range is formed by upthrusting of the crust, its rocks are subject to erosion by water and ice. The sharp peaks and serrated edges have little to do with the forces that make the mountains initially. Instead, they result from the processes that tear down mountains. Ice is an especially effective sculptor of rock ([link]). In a world without moving ice or running water (such as the Moon or Mercury), mountains remain smooth and dull.
Mountains on Earth.

The Torres del Paine are a young region of Earth's crust where sharp mountain peaks are being sculpted by glaciers. We owe the beauty of our young, steep mountains to the erosion by ice and water. (credit: David Morrison)

Volcanoes

Volcanoes mark locations where lava rises to the surface. One example is mid ocean ridges, which are long undersea mountain ranges formed by lava rising from Earth's mantle at plate boundaries. A second major kind of volcanic activity is associated with subduction zones, and volcanoes sometimes also appear in regions where continental plates are colliding. In each case, the volcanic activity gives us a way to sample some of the material from deeper within our planet.

Other volcanic activity occurs above mantle "hot spots"—areas far from plate boundaries where heat is nevertheless rising from the interior of Earth.

One of the best-known hot spot is under the island of Hawaii, where it currently supplies the heat to maintain three active volcanoes, two on land and one under the ocean. The Hawaii hot spot has been active for at least 100 million years. As Earth's plates have moved during that time, the hot spot has generated a 3500-kilometer-long chain of volcanic islands. The tallest Hawaiian volcanoes are among the largest individual mountains on Earth, more than 100 kilometers in diameter and rising 9 kilometers above the ocean floor. One of the Hawaiian volcanic mountains, the now-dormant Mauna Kea, has become one of the world's great sites for doing astronomy.

Note:
The US Geological Service provides an interactive map of the famous "ring of fire," which is the chain of volcanoes surrounding the Pacific Ocean, and shows the Hawaiian "hot spot" enclosed within.

Not all volcanic eruptions produce mountains. If lava flows rapidly from long cracks, it can spread out to form lava plains. The largest known terrestrial eruptions, such as those that produced the Snake River basalts in the northwestern United States or the Deccan plains in India, are of this type. Similar lava plains are found on the Moon and the other terrestrial planets.

Key Concepts and Summary

Terrestrial rocks can be classified as igneous, sedimentary, or metamorphic. A fourth type, primitive rock, is not found on Earth. Our planet's geology is dominated by plate tectonics, in which crustal plates move slowly in response to mantle convection. The surface expression of plate tectonics includes continental drift, recycling of the ocean floor, mountain building, rift zones, subduction zones, faults, earthquakes, and volcanic eruptions of lava from the interior.

Glossary

convection
: movement caused within a gas or liquid by the tendency of hotter, and therefore less dense material, to rise and colder, denser material to sink under the influence of gravity, which consequently results in transfer of heat

fault
: in geology, a crack or break in the crust of a planet along which slippage or movement can take place, accompanied by seismic activity

igneous rock
: rock produced by cooling from a molten state

metamorphic rock
: rock produced by physical and chemical alteration (without melting) under high temperature and pressure

plate tectonics
: the motion of segments or plates of the outer layer of a planet over the underlying mantle

primitive rock
: rock that has not experienced great heat or pressure and therefore remains representative of the original condensed materials from the solar nebula

rift zone
: in geology, a place where the crust is being torn apart by internal forces generally associated with the injection of new material from the mantle and with the slow separation of tectonic plates

sedimentary rock
: rock formed by the deposition and cementing of fine grains of material, such as pieces of igneous rock or the shells of living things

subduction
: the sideways and downward movement of the edge of a plate of Earth’s crust into the mantle beneath another plate

volcano
: a place where material from a planet’s mantle erupts on its surface

Earth’s Atmosphere

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Differentiate between Earth’s various atmospheric layers
- Describe the chemical composition and possible origins of our atmosphere
- Explain the difference between weather and climate

We live at the bottom of the ocean of air that envelops our planet. The atmosphere, weighing down upon Earth’s surface under the force of gravity, exerts a pressure at sea level that scientists define as 1 **bar** (a term that comes from the same root as *barometer*, an instrument used to measure atmospheric pressure). A bar of pressure means that each square centimeter of Earth’s surface has a weight equivalent to 1.03 kilograms pressing down on it. Humans have evolved to live at this pressure; make the pressure a lot lower or higher and we do not function well.

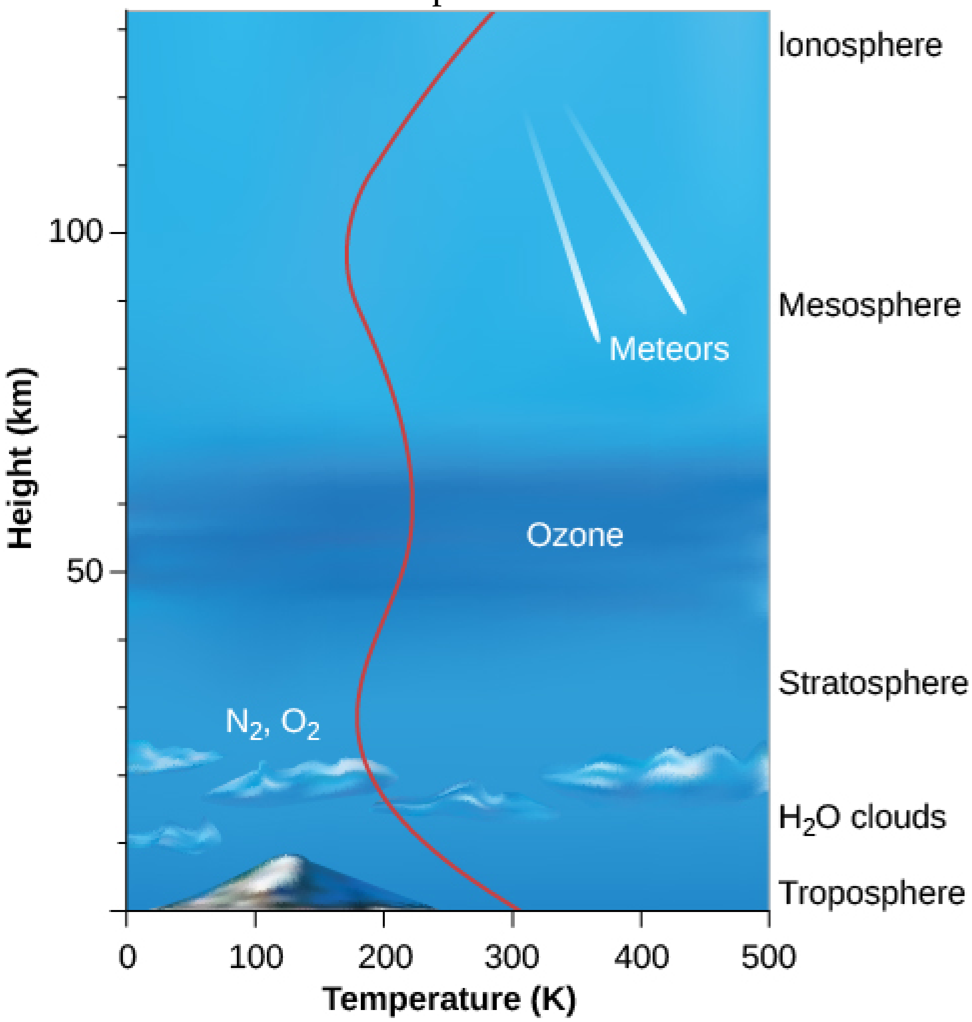
The total mass of Earth’s atmosphere is about 5×10^{18} kilograms. This sounds like a large number, but it is only about a millionth of the total mass of Earth. The atmosphere represents a smaller fraction of Earth than the fraction of your mass represented by the hair on your head.

Structure of the Atmosphere

The structure of the atmosphere is illustrated in [link]. Most of the atmosphere is concentrated near the surface of Earth, within about the

bottom 10 kilometers where clouds form and airplanes fly. Within this region—called the **troposphere**—warm air, heated by the surface, rises and is replaced by descending currents of cooler air; this is an example of convection. This circulation generates clouds and wind. Within the troposphere, temperature decreases rapidly with increasing elevation to values near 50 °C below freezing at its upper boundary, where the **stratosphere** begins. Most of the stratosphere, which extends to about 50 kilometers above the surface, is cold and free of clouds.

Structure of Earth’s Atmosphere.



Height increases up the left side of the diagram, and the names of the different atmospheric layers are shown at the right. In the upper ionosphere, ultraviolet radiation from the Sun can strip electrons from their atoms, leaving the atmosphere ionized. The curving

red line shows the temperature (see the scale on the *x*-axis).

Near the top of the stratosphere is a layer of **ozone** (O_3), a heavy form of oxygen with three atoms per molecule instead of the usual two. Because ozone is a good absorber of ultraviolet light, it protects the surface from some of the Sun's dangerous ultraviolet radiation, making it possible for life to exist on Earth. The breakup of ozone adds heat to the stratosphere, reversing the decreasing temperature trend in the troposphere. Because ozone is essential to our survival, we reacted with justifiable concern to evidence that became clear in the 1980s that atmospheric ozone was being destroyed by human activities. By international agreement, the production of industrial chemicals that cause ozone depletion, called chlorofluorocarbons, or CFCs, has been phased out. As a result, ozone loss has stopped and the "ozone hole" over the Antarctic is shrinking gradually. This is an example of how concerted international action can help maintain the habitability of Earth.

Note:
Visit NASA's scientific visualization studio for a short video of what would have happened to Earth's ozone layer by 2065 if CFCs had not been regulated.

At heights above 100 kilometers, the atmosphere is so thin that orbiting satellites can pass through it with very little friction. Many of the atoms are ionized by the loss of an electron, and this region is often called the ionosphere. At these elevations, individual atoms can occasionally escape completely from the gravitational field of Earth. There is a continuous, slow leaking of atmosphere—especially of lightweight atoms, which move faster than heavy ones. Earth's atmosphere cannot, for example, hold on for long to hydrogen or helium, which escape into space. Earth is not the only planet to experience atmosphere leakage. Atmospheric leakage also created Mars'

thin atmosphere. Venus' dry atmosphere evolved because its proximity to the Sun vaporized and dissociated any water, with the component gases lost to space.

Atmospheric Composition and Origin

At Earth's surface, the atmosphere consists of 78% nitrogen (N_2), 21% oxygen (O_2), and 1% argon (Ar), with traces of water vapor (H_2O), carbon dioxide (CO_2), and other gases. Variable amounts of dust particles and water droplets are also found suspended in the air.

A complete census of Earth's volatile materials, however, should look at more than the gas that is now present. *Volatile* materials are those that evaporate at a relatively low temperature. If Earth were just a little bit warmer, some materials that are now liquid or solid might become part of the atmosphere. Suppose, for example, that our planet were heated to above the boiling point of water (100 °C, or 373 K); that's a large change for humans, but a small change compared to the range of possible temperatures in the universe. At 100 °C, the oceans would boil and the resulting water vapor would become a part of the atmosphere.

To estimate how much water vapor would be released, note that there is enough water to cover the entire Earth to a depth of about 300 meters. Because the pressure exerted by 10 meters of water is equal to about 1 bar, the average pressure at the ocean floor is about 300 bars. Water weighs the same whether in liquid or vapor form, so if the oceans boiled away, the atmospheric pressure of the water would still be 300 bars. Water would therefore greatly dominate Earth's atmosphere, with nitrogen and oxygen reduced to the status of trace constituents.

On a warmer Earth, another source of additional atmosphere would be found in the sedimentary carbonate rocks of the crust. These minerals contain abundant carbon dioxide. If all these rocks were heated, they would release about 70 bars of CO_2, far more than the current CO_2 pressure of only 0.0005 bar. Thus, the atmosphere of a warm Earth would be dominated by water vapor and carbon dioxide, with a surface pressure nearing 400 bars.

Several lines of evidence show that the composition of Earth's atmosphere has changed over our planet's history. Scientists can infer the amount of atmospheric oxygen, for example, by studying the chemistry of minerals that formed at various times. We examine this issue in more detail later in this chapter.

Today we see that CO_2, H_2O, sulfur dioxide (SO_2), and other gases are released from deeper within Earth through the action of volcanoes. (For CO_2, the primary source today is the burning of fossil fuels, which releases far more CO_2 than that from volcanic eruptions.) Much of this apparently new gas, however, is recycled material that has been subducted through plate tectonics. But where did our planet's original atmosphere come from?

Three possibilities exist for the original source of Earth's atmosphere and oceans: (1) the atmosphere could have been formed with the rest of Earth as it accumulated from debris left over from the formation of the Sun; (2) it could have been released from the interior through volcanic activity, subsequent to the formation of Earth; or (3) it may have been derived from impacts by comets and asteroids from the outer parts of the solar system. Current evidence favors a combination of the interior and impact sources.

Weather and Climate

All planets with atmospheres have *weather*, which is the name we give to the circulation of the atmosphere. The energy that powers the weather is derived primarily from the sunlight that heats the surface. Both the rotation of the planet and slower seasonal changes cause variations in the amount of sunlight striking different parts of Earth. The atmosphere and oceans redistribute the heat from warmer to cooler areas. Weather on any planet represents the response of its atmosphere to changing inputs of energy from the Sun (see [link] for a dramatic example).

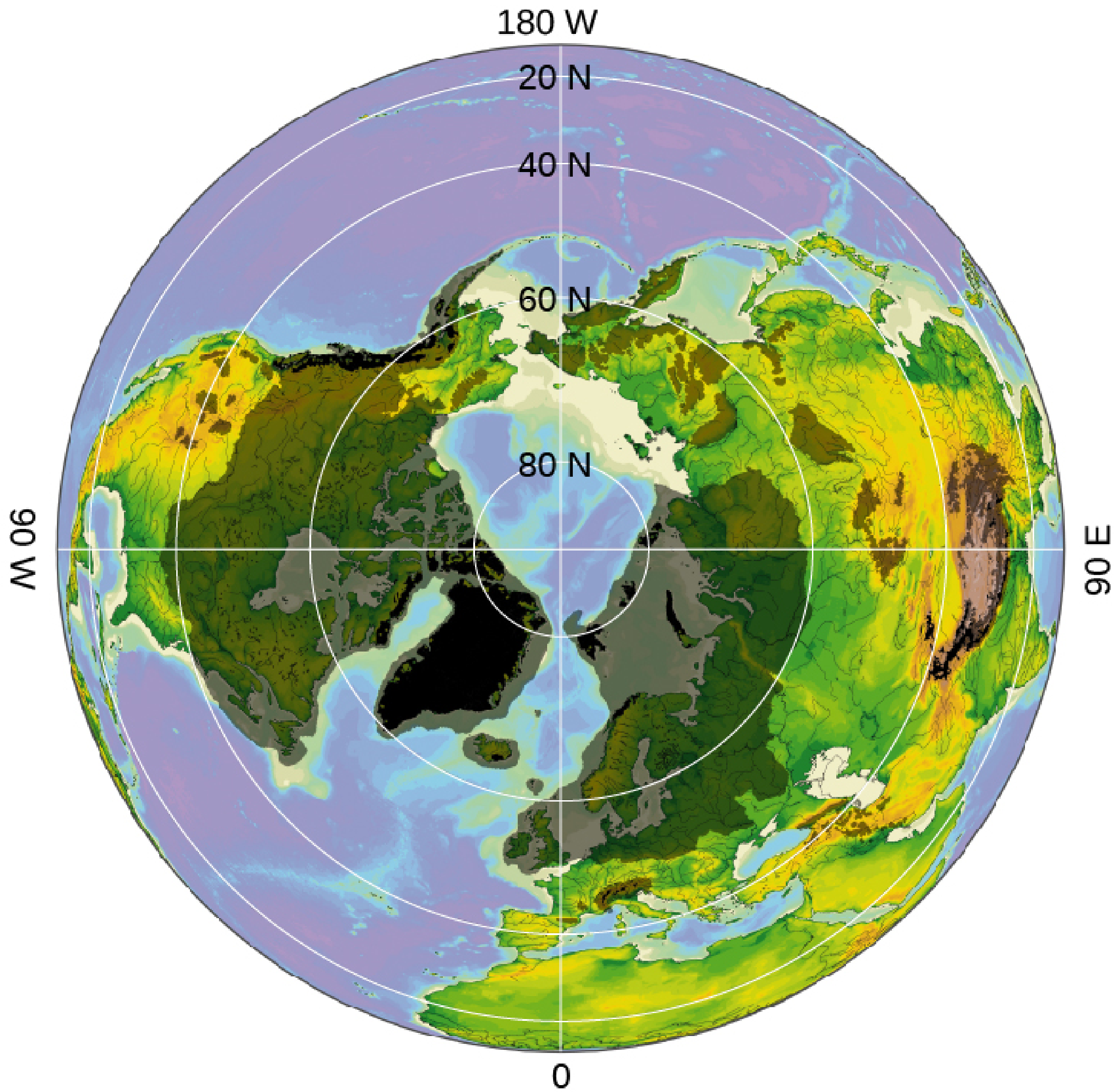
Storm from Space.

This satellite image shows Hurricane Irene in 2011, shortly before the storm hit land in New York City. The combination of Earth's tilted axis of rotation, moderately rapid rotation, and oceans of liquid water can lead to violent weather on our planet. (credit: NASA/NOAA GOES Project)

Climate is a term used to refer to the effects of the atmosphere that last through decades and centuries. Changes in climate (as opposed to the random variations in weather from one year to the next) are often difficult to detect over short time periods, but as they accumulate, their effect can be devastating. One saying is that "Climate is what you expect, and weather is what you get." Modern farming is especially sensitive to temperature and rainfall; for example, calculations indicate that a drop of only 2 °C throughout the growing season would cut the wheat production by half in Canada and the United States. At the other extreme, an increase of 2 °C in

the average temperature of Earth would be enough to melt many glaciers, including much of the ice cover of Greenland, raising sea level by as much as 10 meters, flooding many coastal cities and ports, and putting small islands completely under water.

The best documented changes in Earth's climate are the great ice ages, which have lowered the temperature of the Northern Hemisphere periodically over the past half million years or so ([link]). The last ice age, which ended about 14,000 years ago, lasted some 20,000 years. At its height, the ice was almost 2 kilometers thick over Boston and stretched as far south as New York City.
Ice Age.



This computer-generated image shows the frozen areas of the Northern Hemisphere during past ice ages from the vantage point of looking down on the North Pole. The area in black indicates the most recent glaciation (coverage by glaciers), and the area in gray shows the maximum level of glaciation ever reached. (credit: modification of work by Hannes Grobe/AWI)

These ice ages were primarily the result of changes in the tilt of Earth's rotational axis, produced by the gravitational effects of the other planets. We are less certain about evidence that at least once (and perhaps twice) about a billion years ago, the entire ocean froze over, a situation called *snowball Earth*.

The development and evolution of life on Earth has also produced changes in the composition and temperature of our planet's atmosphere, as we shall see in the next section.

Key Concepts and Summary

The atmosphere has a surface pressure of 1 bar and is composed primarily of N_2 and O_2, plus such important trace gases as H_2O, CO_2, and O_3. Its structure consists of the troposphere, stratosphere, mesosphere, and ionosphere. Changing the composition of the atmosphere also influences the temperature. Atmospheric circulation (weather) is driven by seasonally changing deposition of sunlight. Many longer term climate variations, such as the ice ages, are related to changes in the planet's orbit and axial tilt.

Glossary

bar
: a force of 100,000 Newtons acting on a surface area of 1 square meter; the average pressure of Earth's atmosphere at sea level is 1.013 bars

ozone
: (O_3) a heavy molecule of oxygen that contains three atoms rather than the more normal two

stratosphere
: the layer of Earth's atmosphere above the troposphere and below the ionosphere

troposphere
: the lowest level of Earth's atmosphere, where most weather takes place

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Outline the origins and subsequent diversity of life on Earth
- Explain the ways that life and geological activity have influenced the evolution of the atmosphere
- Describe the causes and effects of the atmospheric greenhouse effect and global warming
- Describe the impact of human activity on our planet's atmosphere and ecology

As far as we know, Earth seems to be the only planet in the solar system with life. The origin and development of life are an important part of our planet's story. Life arose early in Earth's history, and living organisms have been interacting with their environment for billions of years. We recognize that life-forms have evolved to adapt to the environment on Earth, and we are now beginning to realize that Earth itself has been changed in important ways by the presence of living matter. The study of the coevolution of life and our planet is one of the subjects of the modern science of *astrobiology*.

The Origin of Life

The record of the birth of life on Earth has been lost in the restless motions of the crust. According to chemical evidence, by the time the oldest surviving rocks were formed about 3.9 billion years ago, life already existed. At 3.5 billion years ago, life had achieved the sophistication to

build large colonies called *stromatolites*, a form so successful that stromatolites still grow on Earth today ([link]). But, few rocks survive from these ancient times, and abundant fossils have been preserved only during the past 600 million years—less than 15% of our planet's history.

Cross-Sections of Fossil Stromatolites.

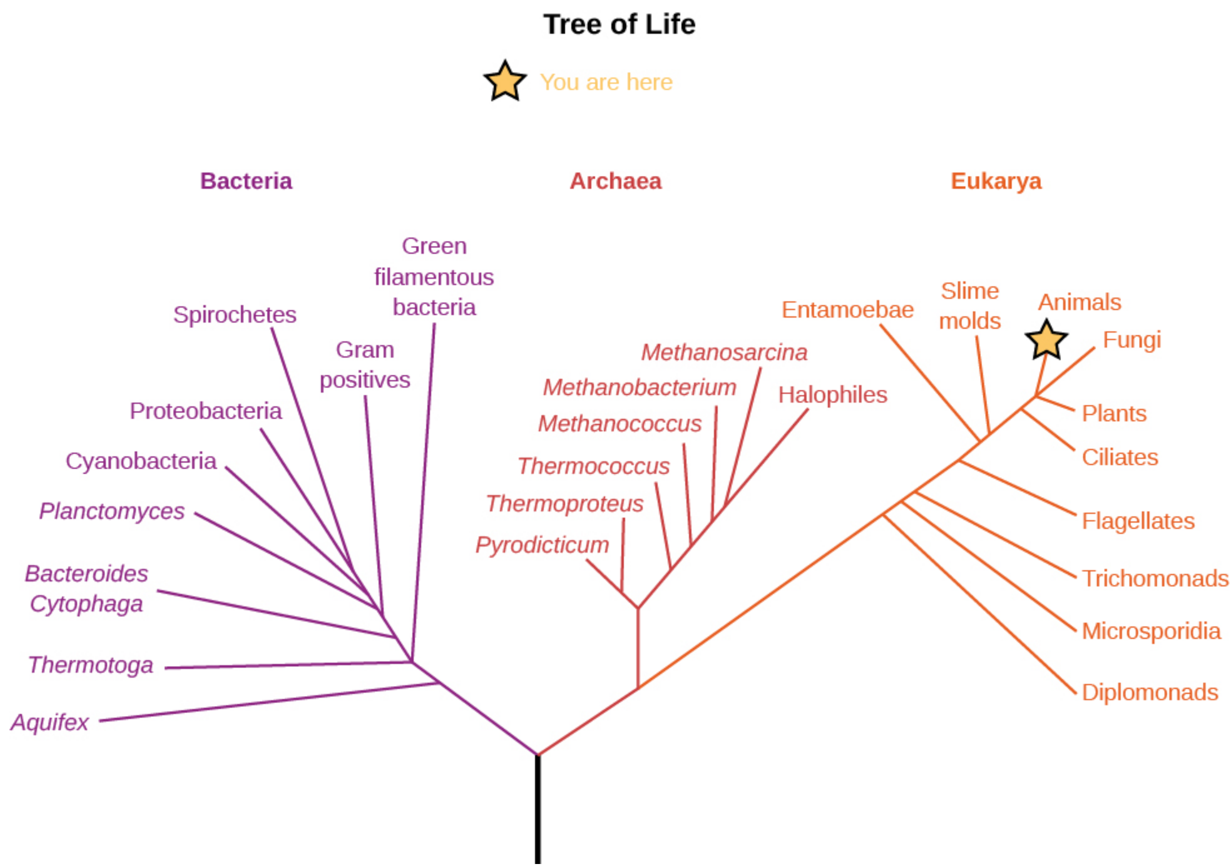
This polished cross-section of a fossilized colony of stromatolites dates to the Precambrian Era. The layered, domelike structures are mats of sediment trapped in shallow waters by large numbers of blue-green bacteria that can photosynthesize. Such colonies of microorganisms date back more than 3 billion years. (credit: James St. John)

There is little direct evidence about the actual origin of life. We know that the atmosphere of early Earth, unlike today's, contained abundant carbon dioxide and some methane, but no oxygen gas. In the absence of oxygen, many complex chemical reactions are possible that lead to the production of amino acids, proteins, and other chemical building blocks of life. Therefore, it seems likely that these chemical building blocks were available very early in Earth's history and they would have combined to make living organisms.

For tens of millions of years after Earth's formation, life (perhaps little more than large molecules, like the viruses of today) probably existed in warm, nutrient-rich seas, living off accumulated organic chemicals. When this easily accessible food became depleted, life began the long evolutionary road that led to the vast numbers of different organisms on Earth today. As it did so, life began to influence the chemical composition of the atmosphere.

In addition to the study of life's history as revealed by chemical and fossil evidence in ancient rocks, scientists use tools from the rapidly advancing fields of genetics and *genomics*—the study of the genetic code that is shared by all life on Earth. While each individual has a unique set of genes (which is why genetic "fingerprinting" is so useful for the study of crime), we also have many genetic traits in common. Your *genome*, the complete map of the DNA in your body, is identical at the 99.9% level to that of Julius Caesar or Marie Curie. At the 99% level, human and chimpanzee genomes are the same. By looking at the gene sequences of many organisms, we can determine that all life on Earth is descended from a common ancestor, and we can use the genetic variations among species as a measure of how closely different species are related.

These genetic analysis tools have allowed scientists to construct what is called the "tree of life" ([link]). This diagram illustrates the way organisms are related by examining one sequence of the nucleic acid RNA that all species have in common. This figure shows that life on Earth is dominated by microscopic creatures that you have probably never heard of. Note that the plant and animal kingdoms are just two little branches at the far right. Most of the diversity of life, and most of our evolution, has taken place at the microbial level. Indeed, it may surprise you to know that there are more microbes in a bucket of soil than there are stars in the Galaxy. You may want to keep this in mind when, later in this book, we turn to the search for life on other worlds. The "aliens" that are most likely to be out there are microbes.
Tree of Life.



This chart shows the main subdivisions of life on Earth and how they are related. Note that the animal and plant kingdoms are just short branches on the far right, along with the fungi. The most fundamental division of Earth's living things is onto three large domains called bacteria, archaea, and eukarya. Most of the species listed are microscopic. (credit: modification of work by Eric Gaba)

Such genetic studies lead to other interesting conclusions as well. For example, it appears that the earliest surviving terrestrial life-forms were all adapted to live at high temperatures. Some biologists think that life might actually have begun in locations on our planet that were extremely hot. Yet another intriguing possibility is that life began on Mars (which cooled sooner) rather than Earth and was "seeded" onto our planet by meteorites traveling from Mars to Earth. Mars rocks are still making their way to

Earth, but so far none has shown evidence of serving as a “spaceship” to carry microorganisms from Mars to Earth.

The Evolution of the Atmosphere

One of the key steps in the evolution of life on Earth was the development of blue-green algae, a very successful life-form that takes in carbon dioxide from the environment and releases oxygen as a waste product. These successful microorganisms proliferated, giving rise to all the lifeforms we call plants. Since the energy for making new plant material from chemical building blocks comes from sunlight, we call the process **photosynthesis**.

Studies of the chemistry of ancient rocks show that Earth’s atmosphere lacked abundant free oxygen until about 2 billion years ago, despite the presence of plants releasing oxygen by photosynthesis. Apparently, chemical reactions with Earth’s crust removed the oxygen gas as quickly as it formed. Slowly, however, the increasing evolutionary sophistication of life led to a growth in the plant population and thus increased oxygen production. At the same time, it appears that increased geological activity led to heavy erosion on our planet’s surface. This buried much of the plant carbon before it could recombine with oxygen to form CO_2.

Free oxygen began accumulating in the atmosphere about 2 billion years ago, and the increased amount of this gas led to the formation of Earth’s ozone layer (recall that ozone is a triple molecule of oxygen, O_3), which protects the surface from deadly solar ultraviolet light. Before that, it was unthinkable for life to venture outside the protective oceans, so the landmasses of Earth were barren.

The presence of oxygen, and hence ozone, thus allowed colonization of the land. It also made possible a tremendous proliferation of animals, which lived by taking in and using the organic materials produced by plants as their own energy source.

As animals evolved in an environment increasingly rich in oxygen, they were able to develop techniques for breathing oxygen directly from the atmosphere. We humans take it for granted that plenty of free oxygen is

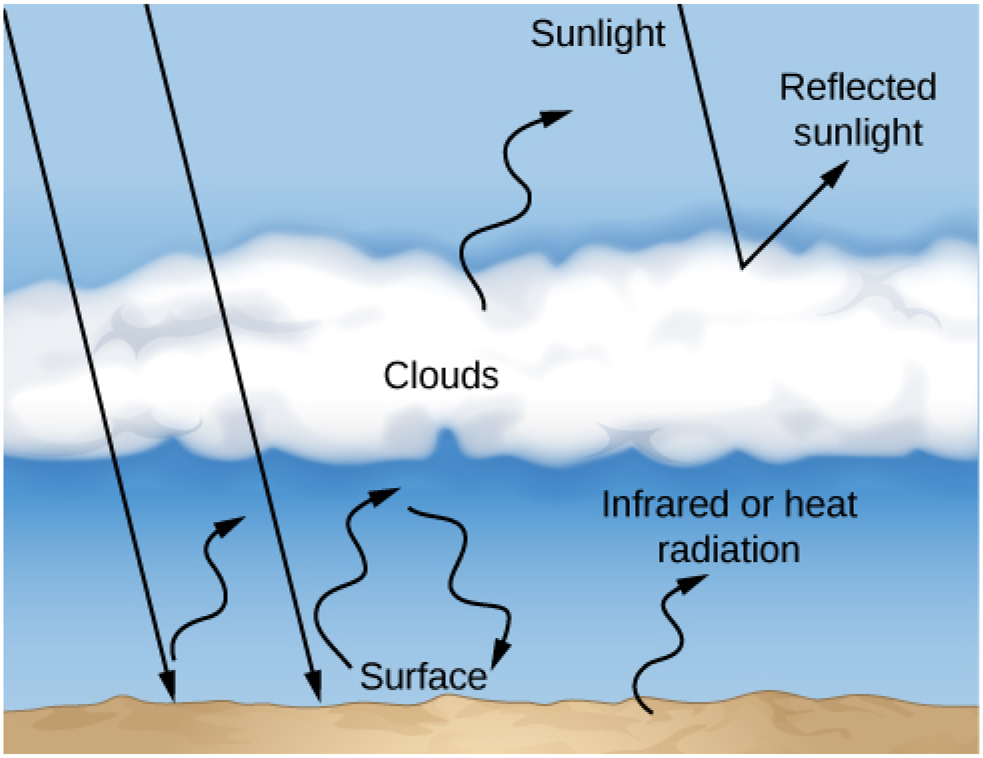
available in Earth’s atmosphere, and we use it to release energy from the food we take in. Although it may seem funny to think of it this way, we are lifeforms that have evolved to breathe in the waste product of plants. It is plants and related microbes that are the primary producers, using sunlight to create energy-rich “food” for the rest of us.

On a planetary scale, one of the consequences of life has been a decrease in atmospheric carbon dioxide. In the absence of life, Earth would probably have an atmosphere dominated by CO_2, like Mars or Venus. But living things, in combination with high levels of geological activity, have effectively stripped our atmosphere of most of this gas.

The Greenhouse Effect and Global Warming

We have a special interest in the carbon dioxide content of the atmosphere because of the key role this gas plays in retaining heat from the Sun through a process called the **greenhouse effect**. To understand how the greenhouse effect works, consider the fate of sunlight that strikes the surface of Earth. The light penetrates our atmosphere, is absorbed by the ground, and heats the surface layers. At the temperature of Earth’s surface, that energy is then reemitted as infrared or heat radiation ([link]). However, the molecules of our atmosphere, which allow visible light through, are good at absorbing infrared energy. As a result, CO_2 (along with methane and water vapor) acts like a blanket, trapping heat in the atmosphere and impeding its flow back to space. To maintain an energy balance, the temperature of the surface and lower atmosphere must increase until the total energy radiated by Earth to space equals the energy received from the Sun. The more CO_2 there is in our atmosphere, the higher the temperature at which Earth’s surface reaches a new balance.

How the Greenhouse Effect Works.



Sunlight that penetrates to Earth's lower atmosphere and surface is reradiated as infrared or heat radiation, which is trapped by greenhouse gases such as water vapor, methane, and CO_2 in the atmosphere. The result is a higher surface temperature for our planet.

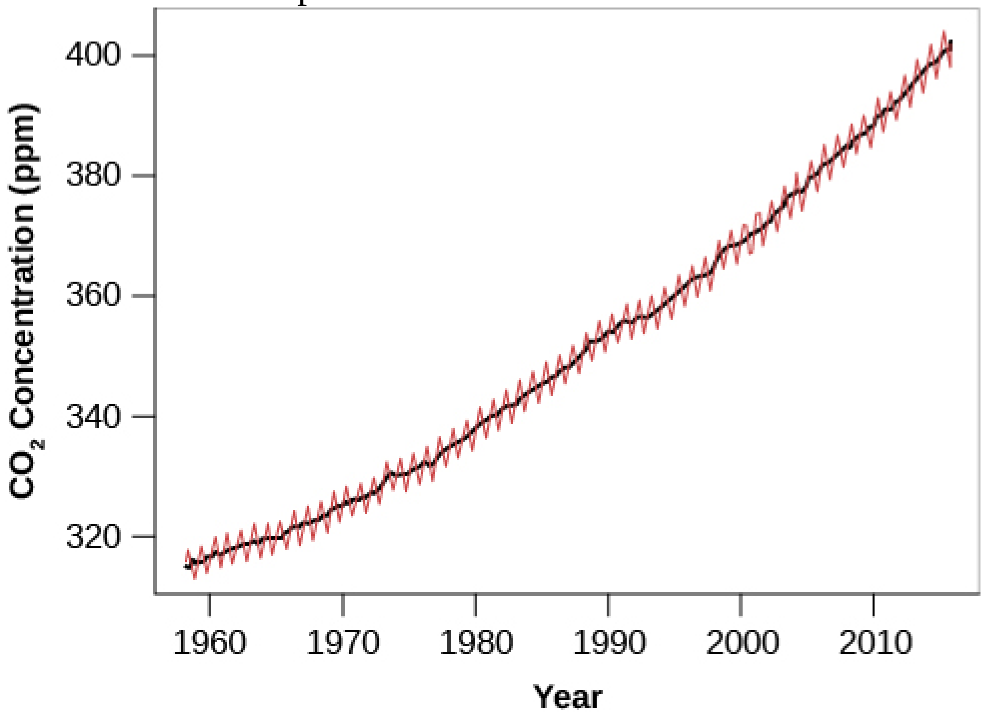
The greenhouse effect in a planetary atmosphere is similar to the heating of a gardener's greenhouse or the inside of a car left out in the Sun with the windows rolled up. In these examples, the window glass plays the role of **greenhouse gases**, letting sunlight in but reducing the outward flow of heat radiation. As a result, a greenhouse or car interior winds up much hotter than would be expected from the heating of sunlight alone. On Earth, the current greenhouse effect elevates the surface temperature by about 23 °C. Without this greenhouse effect, the average surface temperature would be well below freezing and Earth would be locked in a global ice age.

That's the good news; the bad news is that the heating due to the greenhouse effect is increasing. Modern industrial society depends on

energy extracted from burning fossil fuels. In effect, we are exploiting the energy-rich material created by photosynthesis tens of millions of years ago. As these ancient coal and oil deposits are oxidized (burned using oxygen), large quantities of carbon dioxide are released into the atmosphere. The problem is exacerbated by the widespread destruction of tropical forests, which we depend on to extract CO_2 from the atmosphere and replenish our supply of oxygen. In the past century of increased industrial and agricultural development, the amount of CO_2 in the atmosphere increased by about 30% and continues to rise at more than 0.5% per year.

Before the end of the present century, Earth's CO_2 level is predicted to reach twice the value it had before the industrial revolution ([link]). The consequences of such an increase for Earth's surface and atmosphere (and the creatures who live there) are likely to be complex changes in climate, and may be catastrophic for many species. Many groups of scientists are now studying the effects of such global warming with elaborate computer models, and climate change has emerged as the greatest known threat (barring nuclear war) to both industrial civilization and the ecology of our planet.

Increase of Atmospheric Carbon Dioxide over Time.



Scientists expect that the amount of CO_2 will double its preindustrial level before the end of the twenty-first century. Measurements of the isotopic signatures of this added CO_2 demonstrate that it is mostly coming from burning fossil fuels. (credit: modification of work by NOAA)

Note:
This short PBS video explains the physics of the greenhouse effect.

Already climate change is widely apparent. Around the world, temperature records are constantly set and broken; all but one of the hottest recorded years have taken place since 2000. Glaciers are retreating, and the Arctic Sea ice is now much thinner than when it was first explored with nuclear submarines in the 1950s. Rising sea levels (from both melting glaciers and expansion of the water as its temperature rises) pose one of the most immediate threats, and many coastal cities have plans to build dikes or seawalls to hold back the expected flooding. The rate of temperature increase is without historical precedent, and we are rapidly entering "unknown territory" where human activities are leading to the highest temperatures on Earth in more than 50 million years.

Human Impacts on Our Planet

Earth is so large and has been here for so long that some people have trouble accepting that humans are really changing the planet, its atmosphere, and its climate. They are surprised to learn, for example, that the carbon dioxide released from burning fossil fuels is 100 times greater than that emitted by volcanoes. But, the data clearly tell the story that our climate is changing rapidly, and that almost all of the change is a result of human activity.

This is not the first time that humans have altered our environment dramatically. Some of the greatest changes were caused by our ancestors, before the development of modern industrial society. If aliens had visited Earth 50,000 years ago, they would have seen much of the planet supporting large animals of the sort that now survive only in Africa. The plains of Australia were occupied by giant marsupials such as diprododon and zygomaturus (the size of our elephants today), and a species of kangaroo that stood 10 feet high. North America and North Asia hosted mammoths, saber tooth cats, mastodons, giant sloths, and even camels. The Islands of the Pacific teemed with large birds, and vast forests covered what are now the farms of Europe and China. Early human hunters killed many large mammals and marsupials, early farmers cut down most of the forests, and the Polynesian expansion across the Pacific doomed the population of large birds.

An even greater mass extinction is underway as a result of rapid climate change. In recognition of our impact on the environment, scientists have proposed giving a new name to the current epoch, the *anthropocine*, when human activity started to have a significant global impact. Although not an officially approved name, the concept of "anthropocine" is useful for recognizing that we humans now represent the dominant influence on our planet's atmosphere and ecology, for better or for worse.

Key Concepts and Summary

Life originated on Earth at a time when the atmosphere lacked O_2 and consisted mostly of CO_2. Later, photosynthesis gave rise to free oxygen and ozone. Modern genomic analysis lets us see how the wide diversity of species on the planet are related to each other. CO_2 and methane in the atmosphere heat the surface through the greenhouse effect; today, increasing amounts of atmospheric CO_2 are leading to the global warming of our planet.

Glossary

greenhouse gas

a gas in an atmosphere that absorbs and emits radiation within the thermal infrared range; on Earth, these atmospheric gases primarily include carbon dioxide, methane, and water vapor

greenhouse effect
: the blanketing (absorption) of infrared radiation near the surface of a planet—for example, by CO_2 in its atmosphere

photosynthesis
: a complex sequence of chemical reactions through which some living things can use sunlight to manufacture products that store energy (such as carbohydrates), releasing oxygen as one by-product

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the scarcity of impact craters on Earth compared with other planets and moons
- Describe the evidence for recent impacts on Earth
- Detail how a massive impact changed the conditions for life on Earth, leading to the extinction of the dinosaurs
- Describe how impacts have influenced the evolution of life on Earth
- Discuss the search for objects that could potentially collide with our planet

In discussing Earth's geology earlier in this chapter, we dealt only with the effects of internal forces, expressed through the processes of plate tectonics and volcanism. On the Moon, in contrast, we see primarily craters, produced by the impacts of interplanetary debris such as asteroids and comets. Why don't we see more evidence on Earth of the kinds of impact craters that are so prominent on the Moon and other worlds?

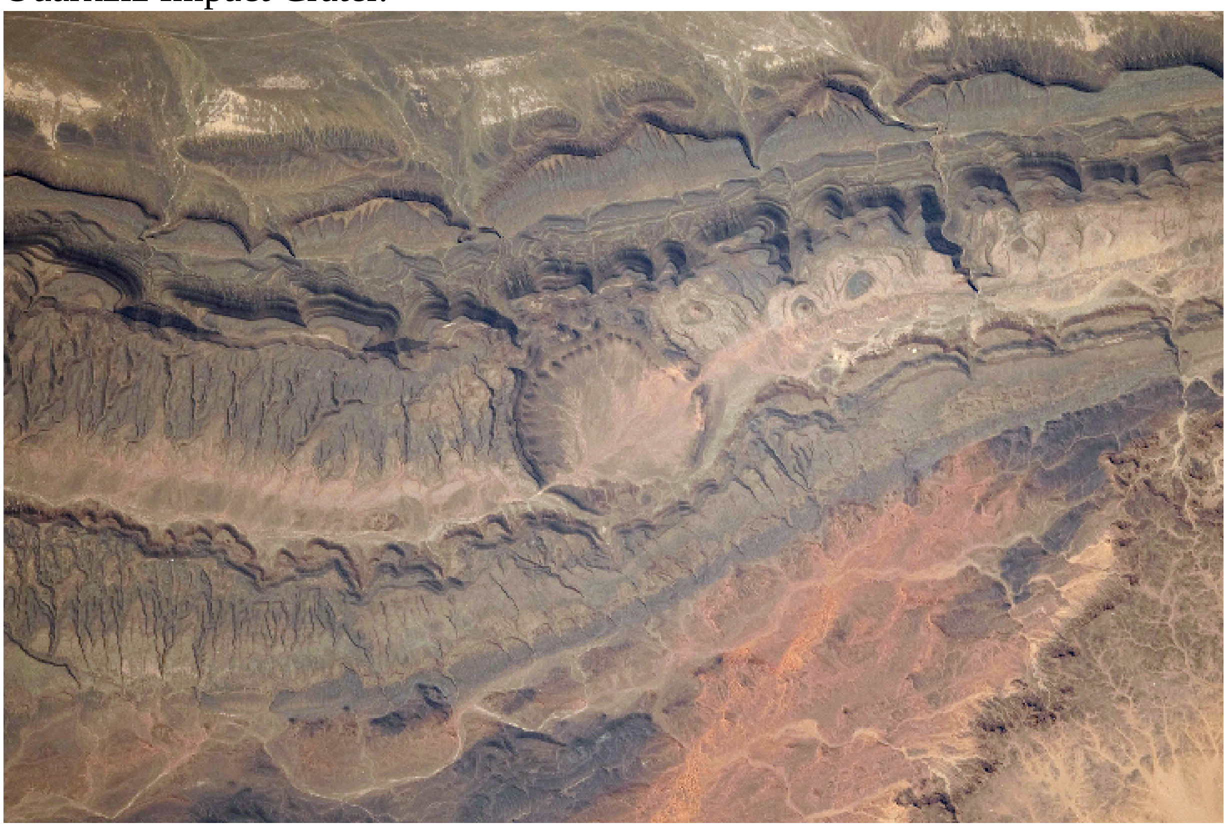
Where Are the Craters on Earth?

It is not possible that Earth escaped being struck by the interplanetary debris that has pockmarked the Moon. From a cosmic perspective, the Moon is almost next door. Our atmosphere does make small pieces of cosmic debris burn up (which we see as *meteors*—commonly called

shooting stars). But, the layers of our air provide no shield against the large impacts that form craters several kilometers in diameter and are common on the Moon.

In the course of its history, Earth must therefore have been impacted as heavily as the Moon. The difference is that, on Earth, these craters are destroyed by our active geology before they can accumulate. As plate tectonics constantly renews our crust, evidence of past cratering events is slowly erased. Only in the past few decades have geologists succeeded in identifying the eroded remnants of many impact craters ([link]). Even more recent is our realization that, over the history of Earth, these impacts have had an important influence on the evolution of life.

Ouarkziz Impact Crater.



Located in Algeria, this crater (the round feature in the center) is the result of a meteor impact during the Cretaceous period. Although the crater has experienced heavy erosion, this image from the International Space Station shows the circular pattern resulting from impact. (credit: modification of work by NASA)

Recent Impacts

The collision of interplanetary debris with Earth is not a hypothetical idea. Evidence of relatively recent impacts can be found on our planet's surface. One well-studied historic collision took place on June 30, 1908, near the Tunguska River in Siberia. In this desolate region, there was a remarkable explosion in the atmosphere about 8 kilometers above the surface. The shock wave flattened more than a thousand square kilometers of forest ([link]). Herds of reindeer and other animals were killed, and a man at a trading post 80 kilometers from the blast was thrown from his chair and knocked unconscious. The blast wave spread around the world, as recorded by instruments designed to measure changes in atmospheric pressure.

Aftermath of the Tunguska Explosion.

This photograph, taken 21 years after the blast, shows a part of the forest that was destroyed by the 5-megaton explosion, resulting when a stony projectile about the size of a small office building (40 meters in diameter) collided with our planet. (credit: modification of work by Leonid Kulik)

Despite this violence, no craters were formed by the Tunguska explosion. Shattered by atmospheric pressure, the stony projectile with a mass of approximately 10,000 tons disintegrated above our planet's surface to create a blast equivalent to a 5-megaton nuclear bomb. Had it been smaller or more fragile, the impacting body would have dissipated its energy at high altitude and probably attracted no attention. Today, such high-altitude atmospheric explosions are monitored regularly by military surveillance systems.

If it had been larger or made of stronger material (such as metal), the Tunguska projectile would have penetrated all the way to the surface of Earth and exploded to form a crater. Instead, only the heat and shock of the atmospheric explosion reached the surface, but the devastation it left behind in Siberia bore witness to the power of such impacts. Imagine if the same rocky impactor had exploded over New York City in 1908; history books might today record it as one of the most deadly events in human history.

Tens of thousands of people witnessed directly the explosion of a smaller (20-meter) projectile over the Russian city of Chelyabinsk on an early winter morning in 2013. It exploded at a height of 21 kilometers in a burst of light brighter than the Sun, and the shockwave of the 0.5-megaton explosion broke tens of thousands of windows and sent hundreds of people to the hospital. Rock fragments (meteorites) were easily collected by people in the area after the blast because they landed on fresh snow.

Note:

Dr. David Morrison, one of the original authors of this textbook, provides a nontechnical talk about the Chelyabinsk explosion, and impacts in general.

The best-known recent crater on Earth was formed about 50,000 years ago in Arizona. The projectile in this case was a lump of iron about 40 meters in diameter. Now called *Meteor Crater* and a major tourist attraction on the way to the Grand Canyon, the crater is about a mile across and has all the features associated with similar-size lunar impact craters ([link]). Meteor Crater is one of the few impact features on Earth that remains relatively intact; some older craters are so eroded that only a trained eye can distinguish them. Nevertheless, more than 150 have been identified. (See the list of suggested online sites at the end of this chapter if you want to find out more about these other impact scars.)

Meteor Crater in Arizona.

Here we see a 50,000-year-old impact crater made by the collision of a 40-meter lump of iron with our planet. Although impact craters are common on less active bodies such as the Moon, this is one of the very few well-preserved craters on Earth. (modification of work by D. Roddy/USGS)

Mass Extinction

The impact that produced Meteor Crater would have been dramatic indeed to any humans who witnessed it (from a safe distance) since the energy release was equivalent to a 10-megaton nuclear bomb. But such explosions are devastating only in their local areas; they have no *global* consequences. Much larger (and rarer) impacts, however, can disturb the ecological balance of the entire planet and thus influence the course of evolution.

The best-documented large impact took place 65 million years ago, at the end of what is now called the Cretaceous period of geological history. This time in the history of life on Earth was marked by a **mass extinction**, in which more than half of the species on our planet died out. There are a dozen or more mass extinctions in the geological record, but this particular event (nicknamed the "great dying") has always intrigued paleontologists because it marks the end of the dinosaur age. For tens of millions of years these great creatures had flourished and dominated. Then, they suddenly disappeared (along with many other species), and thereafter mammals began the development and diversification that ultimately led to all of us.

The object that collided with Earth at the end of the Cretaceous period struck a shallow sea in what is now the Yucatán peninsula of Mexico. Its mass must have been more than a trillion tons, determined from study of a worldwide layer of sediment deposited from the dust cloud that enveloped the planet after its impact. First identified in 1979, this sediment layer is rich in the rare metal iridium and other elements that are relatively abundant in asteroids and comets, but exceedingly rare in Earth's crust. Even though

it was diluted by the material that the explosion excavated from the surface of Earth, this cosmic component can still be identified. In addition, this layer of sediment contains many minerals characteristic of the temperatures and pressures of a gigantic explosion.

The impact that led to the extinction of dinosaurs released energy equivalent to 5 billion Hiroshima-size nuclear bombs and excavated a crater 200 kilometers across and deep enough to penetrate through Earth’s crust. This large crater, named Chicxulub for a small town near its center, has subsequently been buried in sediment, but its outlines can still be identified ([link]). The explosion that created the Chicxulub crater lifted about 100 trillion tons of dust into the atmosphere. We can determine this amount by measuring the thickness of the sediment layer that formed when this dust settled to the surface.

Site of the Chicxulub Crater.

This map shows the location of the impact crater created 65 million

years ago on Mexico’s Yucatán peninsula. The crater is now buried under more than 500 meters of sediment. (credit: modification of work by “Carport”/Wikimedia)

Such a quantity of airborne material would have blocked sunlight completely, plunging Earth into a period of cold and darkness that lasted several months. Many plants dependent on sunlight would have died, leaving plant-eating animals without a food supply. Other worldwide effects included large-scale fires (started by the hot, flying debris from the explosion) that destroyed much of the planet’s forests and grasslands, and a long period in which rainwater around the globe was acidic. It was these environmental effects, rather than the explosion itself, that were responsible for the mass extinction, including the demise of the dinosaurs.

Impacts and the Evolution of Life

It is becoming clear that many—perhaps most—mass extinctions in Earth’s long history resulted from a variety of other causes, but in the case of the dinosaur killer, the cosmic impact certainly played a critical role and may have been the “final straw” in a series of climactic disturbances that resulted in the “great dying.”

A catastrophe for one group of living things, however, may create opportunities for another group. Following each mass extinction, there is a sudden evolutionary burst as new species develop to fill the ecological niches opened by the event. Sixty-five million years ago, our ancestors, the mammals, began to thrive when so many other species died out. We are the lucky beneficiaries of this process.

Impacts by comets and asteroids represent the only mechanisms we know of that could cause truly global catastrophes and seriously influence the evolution of life all over the planet. As paleontologist Stephen Jay Gould of Harvard noted, such a perspective changes fundamentally our view of biological evolution. The central issues for the survival of a species must now include more than just its success in competing with other species and

adapting to slowly changing environments, as envisioned by Darwin's idea of natural selection. Also required is an ability to survive random global catastrophes due to impacts.

Still earlier in its history, Earth was subject to even larger impacts from the leftover debris of planet formation. We know that the Moon was struck repeatedly by objects larger than 100 kilometers in diameter—1000 times more massive than the object that wiped out most terrestrial life 65 million years ago. Earth must have experienced similar large impacts during its first 700 million years of existence. Some of them were probably violent enough to strip the planet of most its atmosphere and to boil away its oceans. Such events would sterilize the planet, destroying any life that had begun. Life may have formed and been wiped out several times before our own microbial ancestors took hold sometime about 4 billion years ago.

The fact that the oldest surviving microbes on Earth are thermophiles (adapted to very high temperatures) can also be explained by such large impacts. An impact that was just a bit too small to sterilize the planet would still have destroyed anything that lived in what we consider "normal" environments, and only the creatures adapted to high temperatures would survive. Thus, the oldest surviving terrestrial lifeforms are probably the remnants of a sort of evolutionary bottleneck caused by repeated large impacts early in the planet's history.

Impacts in Our Future?

The impacts by asteroids and comets that have had such a major influence on life are not necessarily a thing of the past. In the full scope of planetary history, 65 million years ago was just yesterday. Earth actually orbits the Sun within a sort of cosmic shooting gallery, and although major impacts are rare, they are by no means over. Humanity could suffer the same fate as the dinosaurs, or lose a city to the much more frequent impacts like the one over Tunguska, unless we figure out a way to predict the next big impact and to protect our planet. The fact that our solar system is home to some very large planets in outer orbits may be beneficial to us; the gravitational fields of those planets can be very effective at pulling in cosmic debris and shielding us from larger, more frequent impacts.

Beginning in the 1990s, a few astronomers began to analyze the cosmic impact hazard and to persuade the government to support a search for potentially hazardous asteroids. Several small but sophisticated wide-field telescopes are now used for this search, which is called the NASA Spaceguard Survey. Already we know that there are currently no asteroids on a collision course with Earth that are as big (10–15 kilometers) as the one that killed the dinosaurs. The Spaceguard Survey now concentrates on finding smaller potential impactors. By 2015, the search had netted more than 15,000 near-Earth-asteroids, including most of those larger than 1 kilometer. None of those discovered so far poses any danger to us. Of course, we cannot make a similar statement about the asteroids that have not yet been discovered, but these will be found and evaluated one by one for their potential hazard. These asteroid surveys are one of the few really life-and-death projects carried out by astronomers, with a potential to help to save our planet from future major impacts.

Note:
The Torino Impact Hazard Scale is a method for categorizing the impact hazard associated with near-Earth objects such as asteroids and comets. It is a communication tool for astronomers and the public to assess the seriousness of collision predictions by combining probability statistics and known kinetic damage potentials into a single threat value.
Purdue University's "Impact: Earth" calculator lets you input the characteristics of an approaching asteroid to determine the effect of its impact on our planet.

Key Concepts and Summary

Earth, like the Moon and other planets, has been influenced by the impacts of cosmic debris, including such recent examples as Meteor Crater and the Tunguska explosion. Larger past impacts are implicated in some mass extinctions, including the large impact 65 million years ago at the end of the Cretaceous period that wiped out the dinosaurs and many other species. Today, astronomers are working to predict the next impact in advance,

while other scientists are coming to grips with the effect of impacts on the evolution and diversity of life on Earth.

For Further Exploration

Articles

Earth

Collins, W., et al. “The Physical Science behind Climate Change.” *Scientific American* (August 2007): 64. Why scientists are now confident that human activities are changing our planet’s climate.

Glatzmaier, G., & Olson, P. “Probing the Geodynamo.” *Scientific American* (April 2005): 50. Experiments and modeling that tell us about the source and reversals of Earth’s magnetic field.

Gurnis, M. “Sculpting the Earth from Inside Out.” *Scientific American* (March 2001): 40. On motions that lift and lower the continents.

Hartmann, W. “Piecing Together Earth’s Early History.” *Astronomy* (June 1989): 24.

Jewitt, D., & Young, E. “Oceans from the Skies.” *Scientific American* (March 2015): 36. How did Earth get its water after its initial hot period?

Impacts

Boslaugh, M. “In Search of Death-Plunge Asteroids.” *Astronomy (*July 2015): 28. On existing and proposed programs to search for earth-crossing asteroids.

Brusatte, S. “What Killed the Dinosaurs?” *Scientific American* (December 2015): 54. The asteroid hit Earth at an already vulnerable time.

Chyba, C. “Death from the Sky: Tunguska.” *Astronomy* (December 1993): 38. Excellent review article.

Durda, D. “The Chelyabinsk Super-Meteor.” *Sky & Telescope* (June 2013): 24. A nice summary with photos and eyewitness reporting.

Gasperini, L., et al. “The Tunguska Mystery.” *Scientific American* (June 2008): 80. A more detailed exploration of the site of the 1908 impact over Siberia.

Kring, D. “Blast from the Past.” *Astronomy* (August 2006): 46. Six-page introduction to Arizona’s meteor crater.

Websites

Earth

Astronaut Photography of Earth from Space: http://earth.jsc.nasa.gov/. A site with many images and good information.

Exploration of the Earth’s Magnetosphere: http://phy6.org/Education/Intro.html. An educational website by Dr. Daniel Stern.

NASA Goddard: Earth from Space: Fifteen Amazing Things in 15 Years: https://www.nasa.gov/content/goddard/earth-from-space-15-amazing-things-in-15-years. Images and videos that reveal things about our planet and its atmosphere.

U.S. Geological Survey: Earthquake Information Center: http://earthquake.usgs.gov/learn/

Views of the Solar System: http://www.solarviews.com/eng/earth.htm. Overview of Earth.

Impacts

B612 Foundation : https://b612foundation.org/. Set up by several astronauts for research and education about the asteroid threat to Earth and to build a telescope in space to search for dangerous asteroids.

Lunar and Planetary Institute: Introduction to Terrestrial Impact Craters: http://www.lpi.usra.edu/publications/slidesets/craters/. Includes images.

Meteor Crater Tourist Site: http://meteorcrater.com/.

NASA/Jet Propulsion Lab Near Earth Object Program: http://neo.jpl.nasa.gov/neo/.

What Are Near-Earth-Objects: http://spaceguardcentre.com/what-are-neos/. From the British Spaceguard Centre.

Videos

Earth

All Alone in the Night: http://apod.nasa.gov/apod/ap120305.html. Flying over Earth at night (2:30).

Earth Globes Movies (including Earth at night): http://astro.uchicago.edu/cosmus/projects/earth/.

Earth: The Operator's Manual: http://earththeoperatorsmanual.com/feature-video/earth-the-operators-manual. A National Science Foundation–sponsored miniseries on climate change and energy, with geologist Richard Alley (53:43).

PBS NOVA Videos about Earth: http://www.pbs.org/wgbh/nova/earth/. Programs and information about planet Earth. Click full episodes on the menu at left to be taken to a nice array of videos.

U. S. National Weather Service: http://earth.nullschool.net. Real Time Globe of Earth showing wind patterns which can be zoomed and moved to your preferred view.

Impacts

Chelyabinsk Meteor: Can We Survive a Bigger Impact?: https://www.youtube.com/watch?v=Y-e6xyUZLLs. Talk by Dr. David Morrison (1:34:48).

Large Asteroid Impact Simulation: https://www.youtube.com/watch?v=bU1QPtOZQZU. Large asteroid impact simulation from the Discovery Channel (4:45).

Meteor Hits Russia February 15, 2013: https://www.youtube.com/watch?v=dpmXyJrs7iU. Archive of eyewitness footage (10:11).

Sentinel Mission: Finding an Asteroid Headed for Earth: https://www.youtube.com/watch?v=efz8c3ijD_A. Public lecture by astronaut Ed Lu (1:08:57).

Collaborative Group Activities

A. If we can predict that lots of ground movement takes place along subduction zones and faults, then why do so many people live there? Should we try to do anything to discourage people from living in these areas? What inducement would your group offer people to move? Who would pay for the relocation? (Note that two of the original authors of this book live quite close to the San Andreas and Hayward faults. If they wrote this chapter and haven't moved, what are the chances others living in these kinds of areas will move?)
B. After your group reads the feature box on Alfred Wegener: Catching the Drift of Plate Tectonics, discuss some reasons his idea did not catch on right away among scientists. From your studies in this course and in other science courses (in college and before), can you cite other scientific ideas that we now accept but that had controversial

beginnings? Can you think of any scientific theories that are still controversial today? If your group comes up with some, discuss ways scientists could decide whether each theory on your list is right.

C. Suppose we knew that a large chunk of rock or ice (about the same size as the one that hit 65 million years ago) will impact Earth in about 5 years. What could or should we do about it? (The film *Deep Impact* dealt with this theme.) Does your group think that the world as a whole should spend more money to find and predict the orbits of cosmic debris near Earth?
D. Carl Sagan pointed out that any defensive weapon that we might come up with to deflect an asteroid *away* from Earth could be used as an offensive weapon by an unstable dictator in the future to cause an asteroid not heading our way to come toward Earth. The history of human behavior, he noted, has shown that most weapons that are built (even with the best of motives) seem to wind up being used. Bearing this in mind, does your group think we should be building weapons to protect Earth from asteroid or comet impact? Can we afford not to build them? How can we safeguard against these collisions?
E. Is there evidence of climate change in your area over the past century? How would you distinguish a true climate change from the random variations in weather that take place from one year to the next?

Review Questions

Exercise:

Problem: What is the thickest interior layer of Earth? The thinnest?

Exercise:

Problem:

What are Earth's core and mantle made of? Explain how we know.

Exercise:

Problem:

Describe the differences among primitive, igneous, sedimentary, and metamorphic rock, and relate these differences to their origins.

Exercise:

Problem:

Explain briefly how the following phenomena happen on Earth, relating your answers to the theory of plate tectonics

A. earthquakes
B. continental drift
C. mountain building
D. volcanic eruptions
E. creation of the Hawaiian island chain

Exercise:

Problem: What is the source of Earth's magnetic field?

Exercise:

Problem:

Why is the shape of the magnetosphere not spherical like the shape of Earth?

Exercise:

Problem:

Although he did not present a mechanism, what were the key points of Alfred Wegener's proposal for the concept of continental drift?

Exercise:

Problem:

List the possible interactions between Earth's crustal plates that can occur at their boundaries.

Exercise:

Problem:

List, in order of decreasing altitude, the principle layers of Earth's atmosphere.

Exercise:

Problem:

In which atmospheric layer are almost all water-based clouds formed?

Exercise:

Problem:

What is, by far, the most abundant component of Earth's atmosphere?

Exercise:

Problem: In which domain of living things do you find humankind?

Exercise:

Problem:

Describe three ways in which the presence of life has affected the composition of Earth's atmosphere.

Exercise:

Problem: Briefly describe the greenhouse effect.

Exercise:

Problem:

How do impacts by comets and asteroids influence Earth's geology, its atmosphere, and the evolution of life?

Exercise:

Problem:

Why are there so many impact craters on our neighbor world, the Moon, and so few on Earth?

Exercise:

Problem:

Detail some of the anthropogenic changes to Earth's climate and their potential impact on life.

Thought Questions

Exercise:

Problem:

If you wanted to live where the chances of a destructive earthquake were small, would you pick a location near a fault zone, near a mid ocean ridge, near a subduction zone, or on a volcanic island such as Hawaii? What are the relative risks of earthquakes at each of these locations?

Exercise:

Problem:

Which type of object would likely cause more damage if it struck near an urban area: a small metallic object or a large stony/icy one?

Exercise:

Problem:

If all life were destroyed on Earth by a large impact, would new life eventually form to take its place? Explain how conditions would have to change for life to start again on our planet.

Exercise:

Problem: Why is a decrease in Earth's ozone harmful to life?

Exercise:

Problem:

Why are we concerned about the increases in CO_2 and other gases that cause the greenhouse effect in Earth's atmosphere? What steps can we take in the future to reduce the levels of CO_2 in our atmosphere? What factors stand in the way of taking the steps you suggest? (You may include technological, economic, and political factors in your answer.)

Exercise:

Problem:

Do you think scientists should make plans to defend Earth from future asteroid impacts? Is it right to intervene in the same evolutionary process that made the development of mammals (including us) possible after the big impact 65 million years ago?

Figuring for Yourself

Exercise:

Problem:

Europe and North America are moving apart by about 5 m per century. As the continents separate, new ocean floor is created along the mid-Atlantic Rift. If the rift is 5000 km long, what is the total area of new ocean floor created in the Atlantic each century? (Remember that 1 km = 1000 m.)

Exercise:

Problem:

Over the entire Earth, there are 60,000 km of active rift zones, with average separation rates of 5 m/century. How much area of new ocean crust is created each year over the entire planet? (This area is approximately equal to the amount of ocean crust that is subducted since the total area of the oceans remains about the same.)

Exercise:

Problem:

With the information from [link], you can calculate the average age of the ocean floor. First, find the total area of the ocean floor (equal to about 60% of the surface area of Earth). Then compare this with the area created (or destroyed) each year. The average lifetime is the ratio of these numbers: the total area of ocean crust compared to the amount created (or destroyed) each year.

Exercise:

Problem:

What is the volume of new oceanic basalt added to Earth's crust each year? Assume that the thickness of the new crust is 5 km, that there are 60,000 km of rifts, and that the average speed of plate motion is 4 cm/y. What fraction of Earth's entire volume does this annual addition of new material represent?

Exercise:

Problem:

Suppose a major impact that produces a mass extinction takes place on Earth once every 5 million years. Suppose further that if such an event occurred today, you and most other humans would be killed (this would be true even if the human species as a whole survived). Such impact events are random, and one could take place at any time. Calculate the probability that such an impact will occur within the next 50 years (within your lifetime).

Exercise:

Problem:

How do the risks of dying from the impact of an asteroid or comet compare with other risks we are concerned about, such as dying in a car accident or from heart disease or some other natural cause? (Hint: To find the annual risk, go to the library or internet and look up the annual number of deaths from a particular cause in a particular country, and then divide by the population of that country.)

Exercise:

Problem: What fraction of Earth's volume is taken up by the core?

Exercise:

Problem:

Approximately what percentage of Earth's radius is represented by the crust?

Exercise:

Problem:

What is the drift rate of the Pacific plate over the Hawaiian hot spot?

Exercise:

Problem:

What is the percent increase of atmospheric CO_2 in the past 20 years?

Exercise:

Problem:

Estimate the mass of the object that formed Meteor Crater in Arizona.

Glossary

mass extinction
: the sudden disappearance in the fossil record of a large number of species of life, to be replaced by fossils of new species in subsequent layers; mass extinctions are indicators of catastrophic changes in the environment, such as might be produced by a large impact on Earth

Thinking Ahead
class="introduction"
Star Colors.

This long time exposure shows the colors of the stars. The circular motion of the stars across the image is provided by Earth’s rotation. The various colors of the stars are caused by their different temperatures. (credit: modification of work by ESO/A.Santerne)

Everything we know about stars—how they are born, what they are made of, how far away they are, how long they live, and how they will die—we learn by decoding the messages contained in the light and radiation that reaches Earth. What questions should we ask, and how do we find the answers?

We can begin our voyage to the stars by looking at the night sky. It is obvious that stars do not all appear equally bright, nor are they all the same color. To understand the stars, we must first determine their basic properties, such as what their temperatures are, how much material they contain (their masses), and how much energy they produce. Since our Sun is a star, of course the same techniques, including spectroscopy, used to study the Sun can be used to find out what stars are like. As we learn more about the stars, we will use these characteristics to begin assembling clues to the main problems we are interested in solving: How do stars form? How long do they survive? What is their ultimate fate?

The Brightness of Stars

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the difference between luminosity and apparent brightness
- Understand how astronomers specify brightness with magnitudes

Luminosity

Perhaps the most important characteristic of a star is its **luminosity**—the total amount of energy at all wavelengths that it emits per second. Earlier, we saw that the Sun puts out a tremendous amount of energy every second. (And there are stars far more luminous than the Sun out there.) To make the comparison among stars easy, astronomers express the luminosity of other stars in terms of the Sun's luminosity. For example, the luminosity of Sirius is about 25 times that of the Sun. We use the symbol L_{Sun} to denote the Sun's luminosity; hence, that of Sirius can be written as 25 L_{Sun}. In a later chapter, we will see that if we can measure how much energy a star emits and we also know its mass, then we can calculate how long it can continue to shine before it exhausts its nuclear energy and begins to die.

Apparent Brightness

Astronomers are careful to distinguish between the luminosity of the star (the total energy output) and the amount of energy that happens to reach our

eyes or a telescope on Earth. Stars are democratic in how they produce radiation; they emit the same amount of energy in every direction in space. Consequently, only a minuscule fraction of the energy given off by a star actually reaches an observer on Earth. We call the amount of a star's energy that reaches a given area (say, one square meter) each second here on Earth its **apparent brightness**. If you look at the night sky, you see a wide range of apparent brightnesses among the stars. Most stars, in fact, are so dim that you need a telescope to detect them.

If all stars were the same luminosity—if they were like standard bulbs with the same light output—we could use the difference in their apparent brightnesses to tell us something we very much want to know: how far away they are. Imagine you are in a big concert hall or ballroom that is dark except for a few dozen 25-watt bulbs placed in fixtures around the walls. Since they are all 25-watt bulbs, their luminosity (energy output) is the same. But from where you are standing in one corner, they do *not* have the same apparent brightness. Those close to you appear brighter (more of their light reaches your eye), whereas those far away appear dimmer (their light has spread out more before reaching you). In this way, you can tell which bulbs are closest to you. In the same way, if all the stars had the same luminosity, we could immediately infer that the brightest-appearing stars were close by and the dimmest-appearing ones were far away.

To pin down this idea more precisely, recall from the Radiation and Spectra chapter that we know exactly how light fades with increasing distance. The energy we receive is inversely proportional to the square of the distance. If, for example, we have two stars of the same luminosity and one is twice as far away as the other, it will look four times dimmer than the closer one. If it is three times farther away, it will look nine (three squared) times dimmer, and so forth.

Alas, the stars do not all have the same luminosity. (Actually, we are pretty glad about that because having many different types of stars makes the universe a much more interesting place.) But this means that if a star looks dim in the sky, we cannot tell whether it appears dim because it has a low luminosity but is relatively nearby, or because it has a high luminosity but is very far away. To measure the luminosities of stars, we must first

compensate for the dimming effects of distance on light, and to do that, we must know how far away they are. Distance is among the most difficult of all astronomical measurements. We will return to how it is determined after we have learned more about the stars. For now, we will describe how astronomers specify the apparent brightness of stars.

The Magnitude Scale

The process of measuring the apparent brightness of stars is called *photometry* (from the Greek *photo* meaning "light" and *–metry* meaning "to measure"). As we saw Observing the Sky: The Birth of Astronomy, astronomical photometry began with Hipparchus. Around 150 B.C.E., he erected an observatory on the island of Rhodes in the Mediterranean. There he prepared a catalog of nearly 1000 stars that included not only their positions but also estimates of their apparent brightnesses.

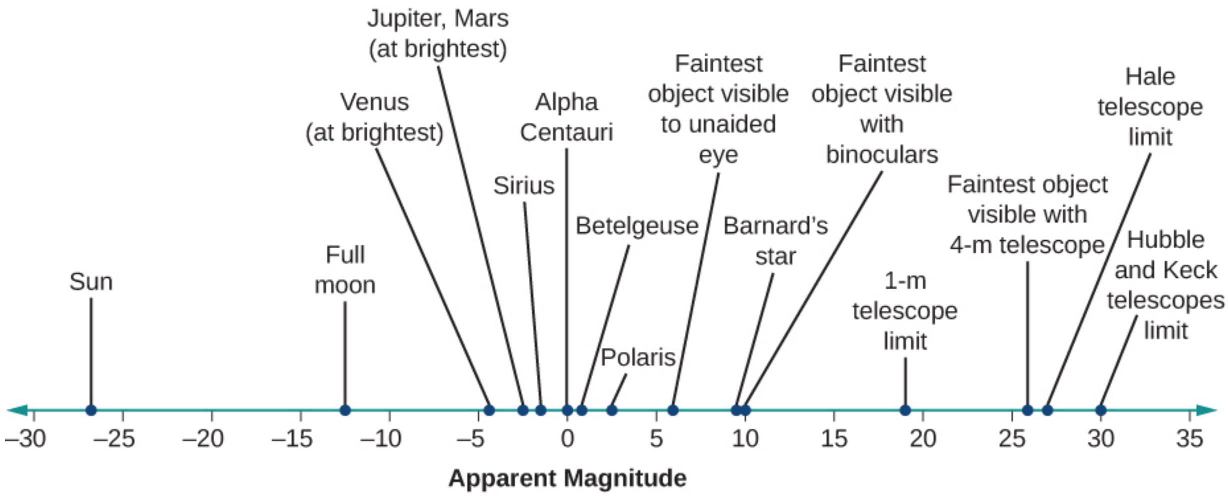
Hipparchus did not have a telescope or any instrument that could measure apparent brightness accurately, so he simply made estimates with his eyes. He sorted the stars into six brightness categories, each of which he called a **magnitude**. He referred to the brightest stars in his catalog as first-magnitudes stars, whereas those so faint he could barely see them were sixth-magnitude stars. During the nineteenth century, astronomers attempted to make the scale more precise by establishing exactly how much the apparent brightness of a sixth-magnitude star differs from that of a first-magnitude star. Measurements showed that we receive about 100 times more light from a first-magnitude star than from a sixth-magnitude star. Based on this measurement, astronomers then defined an accurate magnitude system in which a difference of five magnitudes corresponds exactly to a brightness ratio of 100:1. In addition, the magnitudes of stars are decimalized; for example, a star isn't just a "second-magnitude star," it has a magnitude of 2.0 (or 2.1, 2.3, and so forth). So what number is it that, when multiplied together five times, gives you this factor of 100? Play on your calculator and see if you can get it. The answer turns out to be about 2.5, which is the fifth root of 100. This means that a magnitude 1.0 star and a magnitude 2.0 star differ in brightness by a factor of about 2.5. Likewise, we receive about 2.5 times as much light from a magnitude 2.0 star as from a magnitude 3.0 star. What about the difference between a magnitude 1.0

star and a magnitude 3.0 star? Since the difference is 2.5 times for each "step" of magnitude, the total difference in brightness is $2.5 \times 2.5 = 6.25$ times.

Here are a few rules of thumb that might help those new to this system. If two stars differ by 0.75 magnitudes, they differ by a factor of about 2 in brightness. If they are 2.5 magnitudes apart, they differ in brightness by a factor of 10, and a 4-magnitude difference corresponds to a difference in brightness of a factor of 40.You might be saying to yourself at this point, "Why do astronomers continue to use this complicated system from more than 2000 years ago?" That's an excellent question and, as we shall discuss, astronomers today can use other ways of expressing how bright a star looks. But because this system is still used in many books, star charts, and computer apps, we felt we had to introduce students to it (even though we were very tempted to leave it out.)

The brightest stars, those that were traditionally referred to as first-magnitude stars, actually turned out (when measured accurately) not to be identical in brightness. For example, the brightest star in the sky, Sirius, sends us about 10 times as much light as the average first-magnitude star. On the modern magnitude scale, Sirius, the star with the brightest apparent magnitude, has been assigned a magnitude of −1.5. Other objects in the sky can appear even brighter. Venus at its brightest is of magnitude −4.4, while the Sun has a magnitude of −26.8. [link] shows the range of observed magnitudes from the brightest to the faintest, along with the actual magnitudes of several well-known objects. The important fact to remember when using magnitude is that the system goes backward: the *larger* the magnitude, the *fainter* the object you are observing.

Apparent Magnitudes of Well-Known Objects.



The faintest magnitudes that can be detected by the unaided eye, binoculars, and large telescopes are also shown.

Example:

The Magnitude Equation

Even scientists can't calculate fifth roots in their heads, so astronomers have summarized the above discussion in an equation to help calculate the difference in brightness for stars with different magnitudes. If m_1 and m_2 are the magnitudes of two stars, then we can calculate the ratio of their brightness $\left(\frac{b_2}{b_1}\right)$ using this equation:

Equation:

$$m_1 - m_2 = 2.5 \log\left(\frac{b_2}{b_1}\right) \quad \text{or} \quad \frac{b_2}{b_1} = 2.5^{m_1 - m_2}$$

Here is another way to write this equation:

Equation:

$$\frac{b_2}{b_1} = \left(100^{0.2}\right)^{m_1 - m_2}$$

Let’s do a real example, just to show how this works. Imagine that an astronomer has discovered something special about a dim star (magnitude 8.5), and she wants to tell her students how much dimmer the star is than Sirius. Star 1 in the equation will be our dim star and star 2 will be Sirius.

Solution

Remember, Sirius has a magnitude of −1.5. In that case:

Equation:

$$\begin{aligned}\frac{b_2}{b_1} &= \left(100^{0.2}\right)^{8.5-(-1.5)} = \left(100^{0.2}\right)^{10} \\ &= (100)^2 = 100 \times 100 = 10{,}000\end{aligned}$$

Check Your Learning

It is a common misconception that Polaris (magnitude 2.0) is the brightest star in the sky, but, as we saw, that distinction actually belongs to Sirius (magnitude −1.5). How does Sirius’ apparent brightness compare to that of Polaris?

Note:
Answer:

$$\frac{b_{\text{Sirius}}}{b_{\text{Polaris}}} = \left(100^{0.2}\right)^{2.0-(-1.5)} = \left(100^{0.2}\right)^{3.5} = 100^{0.7} = 25$$

(Hint: If you only have a basic calculator, you may wonder how to take 100 to the 0.7th power. But this is something you can ask Google to do. Google now accepts mathematical questions and will answer them. So try it for yourself. Ask Google, “What is 100 to the 0.7th power?”)

Our calculation shows that Sirius’ apparent brightness is 25 times greater than Polaris’ apparent brightness.

Other Units of Brightness

Although the magnitude scale is still used for visual astronomy, it is not used at all in newer branches of the field. In radio astronomy, for example, no equivalent of the magnitude system has been defined. Rather, radio astronomers measure the amount of energy being collected each second by each square meter of a radio telescope and express the brightness of each source in terms of, for example, watts per square meter.

Similarly, most researchers in the fields of infrared, X-ray, and gamma-ray astronomy use energy per area per second rather than magnitudes to express the results of their measurements. Nevertheless, astronomers in all fields are careful to distinguish between the *luminosity* of the source (even when that luminosity is all in X-rays) and the amount of energy that happens to reach us on Earth. After all, the luminosity is a really important characteristic that tells us a lot about the object in question, whereas the energy that reaches Earth is an accident of cosmic geography.

To make the comparison among stars easy, in this text, we avoid the use of magnitudes as much as possible and will express the luminosity of other stars in terms of the Sun's luminosity. For example, the luminosity of Sirius is 25 times that of the Sun. We use the symbol L_{Sun} to denote the Sun's luminosity; hence, that of Sirius can be written as 25 L_{Sun}.

Key Concepts and Summary

The total energy emitted per second by a star is called its luminosity. How bright a star looks from the perspective of Earth is its apparent brightness. The apparent brightness of a star depends on both its luminosity and its distance from Earth. Thus, the determination of apparent brightness and measurement of the distance to a star provide enough information to calculate its luminosity. The apparent brightnesses of stars are often expressed in terms of magnitudes, which is an old system based on how human vision interprets relative light intensity.

Glossary

apparent brightness

a measure of the amount of light received by Earth from a star or other object—that is, how bright an object appears in the sky, as contrasted with its luminosity

luminosity
: the rate at which a star or other object emits electromagnetic energy into space; the total power output of an object

magnitude
: an older system of measuring the amount of light we receive from a star or other luminous object; the larger the magnitude, the less radiation we receive from the object

Colors of Stars

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Compare the relative temperatures of stars based on their colors
- Understand how astronomers use color indexes to measure the temperatures of stars

Look at the beautiful picture of the stars in the Sagittarius Star Cloud shown in [link]. The stars show a multitude of colors, including red, orange, yellow, white, and blue. As we have seen, stars are not all the same color because they do not all have identical temperatures. To define *color* precisely, astronomers have devised quantitative methods for characterizing the color of a star and then using those colors to determine stellar temperatures. In the chapters that follow, we will provide the temperature of the stars we are describing, and this section tells you how those temperatures are determined from the colors of light the stars give off.

Sagittarius Star Cloud.

This image, which was taken by the Hubble Space Telescope, shows stars in the direction toward the center of the Milky Way Galaxy. The bright stars glitter like colored jewels on a black velvet background. The color of a star indicates its temperature. Blue-white stars are much hotter than the Sun, whereas red stars are cooler. On average, the stars in this field are at a distance of about 25,000 light-years (which means it takes light 25,000 years to traverse the distance from them to us) and the width of the field is about 13.3 light-years. (credit: Hubble Heritage Team (AURA/STScI/NASA))

Color and Temperature

As we learned in The Electromagnetic Spectrum section, Wien's law relates stellar color to stellar temperature. Blue colors dominate the visible light output of very hot stars (with much additional radiation in the ultraviolet). On the other hand, cool stars emit most of their visible light energy at red wavelengths (with more radiation coming off in the infrared) ([link]). The color of a star therefore provides a measure of its intrinsic or true surface temperature (apart from the effects of reddening by interstellar dust, which will be discussed in Between the Stars: Gas and Dust in Space). Color does not depend on the distance to the object. This should be familiar to you from everyday experience. The color of a traffic signal, for example, appears the same no matter how far away it is. If we could somehow take a star, observe it, and then move it much farther away, its apparent brightness (magnitude) would change. But this change in brightness is the same for all wavelengths, and so its color would remain the same.

Example Star Colors and Corresponding Approximate Temperatures

Star Color	Approximate Temperature	Example
Blue	25,000 K	Spica
White	10,000 K	Vega
Yellow	6000 K	Sun
Orange	4000 K	Aldebaran
Red	3000 K	Betelgeuse

Note:
Go to this interactive simulation from the University of Colorado to see the color of a star changing as the temperature is changed.

The hottest stars have temperatures of over 40,000 K, and the coolest stars have temperatures of about 2000 K. Our Sun's surface temperature is about 6000 K; its peak wavelength color is a slightly greenish-yellow. In space, the Sun would look white, shining with about equal amounts of reddish and bluish wavelengths of light. It looks somewhat yellow as seen from Earth's surface because our planet's nitrogen molecules scatter some of the shorter (i.e., blue) wavelengths out of the beams of sunlight that reach us, leaving more long wavelength light behind. This also explains why the sky is blue: the blue sky is sunlight scattered by Earth's atmosphere.

Color Indices

In order to specify the exact color of a star, astronomers normally measure a star's apparent brightness through filters, each of which transmits only the light from a particular narrow band of wavelengths (colors). A crude example of a filter in everyday life is a green-colored, plastic, soft drink bottle, which, when held in front of your eyes, lets only the green colors of light through.

One commonly used set of filters in astronomy measures stellar brightness at three wavelengths corresponding to ultraviolet, blue, and yellow light. The filters are named: U (ultraviolet), B (blue), and V (visual, for yellow). These filters transmit light near the wavelengths of 360 nanometers (nm), 420 nm, and 540 nm, respectively. The brightness measured through each filter is usually expressed in magnitudes. The difference between any two of these magnitudes—say, between the blue and the visual magnitudes (B–V)—is called a **color index**.

Note:

Go to this light and filters simulator for a demonstration of how different light sources and filters can combine to determine the observed spectrum. You can also see how the perceived colors are associated with the spectrum.

By agreement among astronomers, the ultraviolet, blue, and visual magnitudes of the UBV system are adjusted to give a color index of 0 to a star with a surface temperature of about 10,000 K, such as Vega. The B–V color indexes of stars range from −0.4 for the bluest stars, with temperatures of about 40,000 K, to +2.0 for the reddest stars, with temperatures of about 2000 K. The B–V index for the Sun is about +0.65. Note that, by convention, the B–V index is always the “bluer” minus the “redder” color.

Why use a color index if it ultimately implies temperature? Because the brightness of a star through a filter is what astronomers actually measure, and we are always more comfortable when our statements have to do with measurable quantities.

Key Concepts and Summary

Stars have different colors, which are indicators of temperature. The hottest stars tend to appear blue or blue-white, whereas the coolest stars are red. A color index of a star is the difference in the magnitudes measured at any two wavelengths and is one way that astronomers measure and express the temperature of stars.

Glossary

color index
: difference between the magnitudes of a star or other object measured in light of two different spectral regions—for example, blue minus visual (B–V) magnitudes

The Spectra of Stars (and Brown Dwarfs)

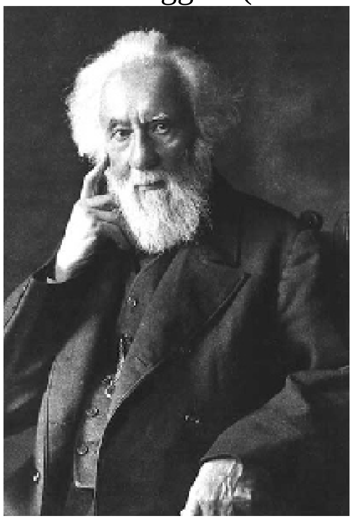
LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe how astronomers use spectral classes to characterize stars
- Explain the difference between a star and a brown dwarf

Measuring colors is only one way of analyzing starlight. Another way is to use a spectrograph to spread out the light into a spectrum (see the Radiation and Spectra and the Astronomical Instruments chapters). In 1814, the German physicist Joseph Fraunhofer observed that the spectrum of the Sun shows dark lines crossing a continuous band of colors. In the 1860s, English astronomers Sir William Huggins and Lady Margaret Huggins ([link]) succeeded in identifying some of the lines in stellar spectra as those of known elements on Earth, showing that the same chemical elements found in the Sun and planets exist in the stars. Since then, astronomers have worked hard to perfect experimental techniques for obtaining and measuring spectra, and they have developed a theoretical understanding of what can be learned from spectra. Today, spectroscopic analysis is one of the cornerstones of astronomical research.

William Huggins (1824–1910) and Margaret Huggins (1848–1915).



William and Margaret Huggins were the first to identify the lines in the spectrum of a star other than

identify the lines in the spectrum of a star other than the Sun; they also took the first spectrogram, or photograph of a stellar spectrum.

Formation of Stellar Spectra

When the spectra of different stars were first observed, astronomers found that they were not all identical. Since the dark lines are produced by the chemical elements present in the stars, astronomers first thought that the spectra differ from one another because stars are not all made of the same chemical elements. This hypothesis turned out to be wrong. *The primary reason that stellar spectra look different is because the stars have different temperatures*. Most stars have nearly the same composition as the Sun, with only a few exceptions.

Hydrogen, for example, is by far the most abundant element in most stars. However, lines of hydrogen are not seen in the spectra of the hottest and the coolest stars. In the atmospheres of the hottest stars, hydrogen atoms are completely ionized. Because the electron and the proton are separated, ionized hydrogen cannot produce absorption lines. (Recall from the Formation of Spectral Lines section, the lines are the result of electrons in orbit around a nucleus changing energy levels.)

In the atmospheres of the coolest stars, hydrogen atoms have their electrons attached and can switch energy levels to produce lines. However, practically all of the hydrogen atoms are in the lowest energy state (unexcited) in these stars and thus can absorb only those photons able to lift an electron from that first energy level to a higher level. Photons with enough energy to do this lie in the ultraviolet part of the electromagnetic spectrum, and there are very few ultraviolet photons in the radiation from a cool star. What this means is that if you observe the spectrum of a very hot or very cool star with a typical telescope on the surface of Earth, the most common element in that star, hydrogen, will show very weak spectral lines or none at all.

The hydrogen lines in the visible part of the spectrum (called *Balmer lines*) are strongest in stars with intermediate temperatures—not too hot and not too cold. Calculations show that the optimum temperature for producing visible hydrogen lines is about 10,000 K. At this temperature, an appreciable number of hydrogen atoms are excited to the second energy level. They can then absorb additional photons, rise to still-higher levels of excitation, and produce a dark absorption line. Similarly, every other chemical element, in each of its possible stages of ionization, has a characteristic temperature at which it is most effective in producing absorption lines in any particular part of the spectrum.

Classification of Stellar Spectra

Astronomers use the patterns of lines observed in stellar spectra to sort stars into a **spectral class**. Because a star's temperature determines which absorption lines are present in its spectrum, these spectral classes are a measure of its surface temperature. There are seven standard spectral classes. From hottest to coldest, these seven spectral classes are designated O, B, A, F, G, K, and M. Recently, astronomers have added three additional classes for even cooler objects—L, T, and Y.

At this point, you may be looking at these letters with wonder and asking yourself why astronomers didn't call the spectral types A, B, C, and so on. You will see, as we tell you the history, that it's an instance where tradition won out over common sense.

In the 1880s, Williamina Fleming devised a system to classify stars based on the strength of hydrogen absorption lines. Spectra with the strongest lines were classified as "A" stars, the next strongest "B," and so on down the alphabet to "O" stars, in which the hydrogen lines were very weak. But we saw above that hydrogen lines alone are not a good indicator for classifying stars, since their lines disappear from the visible light spectrum when the stars get too hot or too cold.

In the 1890s, Annie Jump Cannon revised this classification system, focusing on just a few letters from the original system: A, B, F, G, K, M, and O. Instead of starting over, Cannon also rearranged the existing classes—in order of decreasing temperature—into the sequence we have learned: O, B, A, F, G, K, M. As you can read in the feature on Annie Cannon: Classifier of the Stars in this chapter, she classified around 500,000 stars over her lifetime, classifying up to three stars per minute by looking at the stellar spectra.

Note:
For a deep dive into spectral types, explore the interactive project at the Sloan Digital Sky Survey in which you can practice classifying stars yourself.

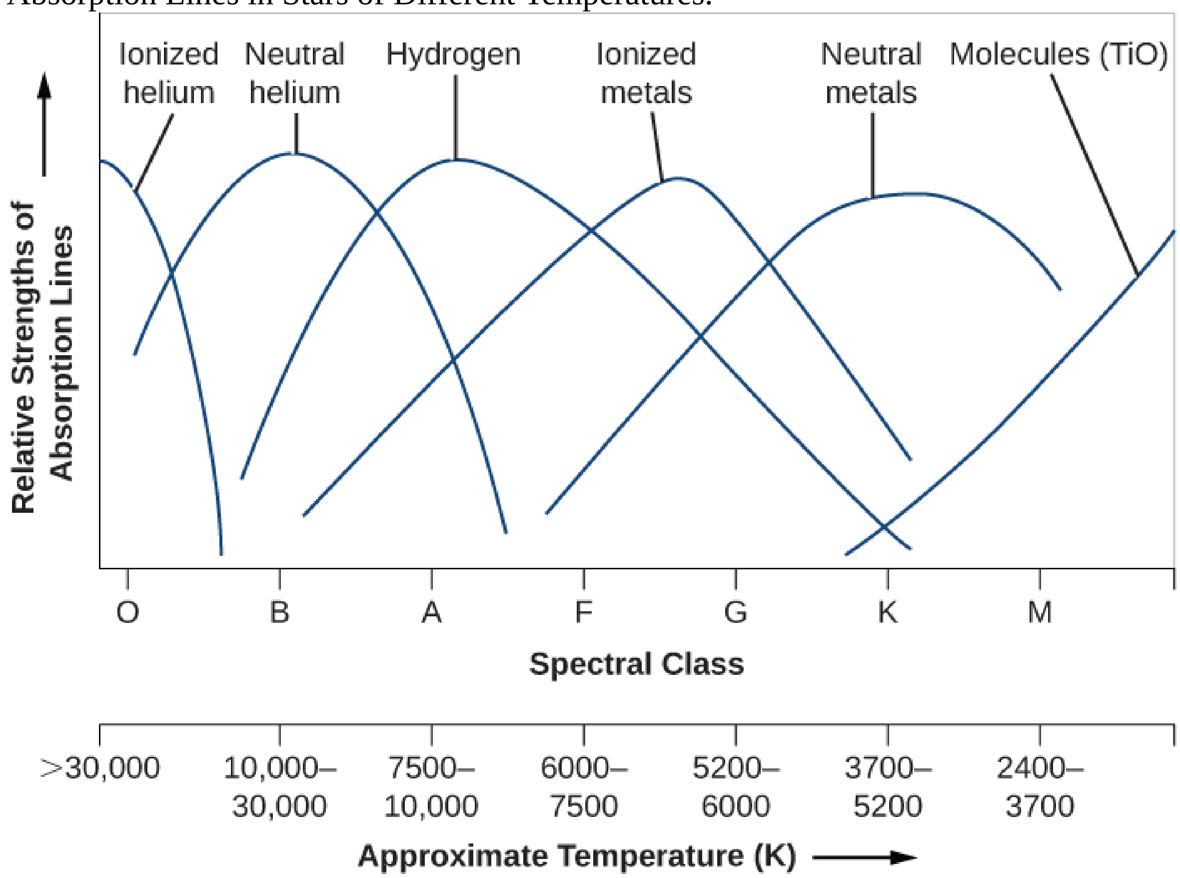
To help astronomers remember this crazy order of letters, Cannon created a mnemonic, "Oh Be A Fine Girl, Kiss Me." (If you prefer, you can easily substitute "Guy" for "Girl.") Other mnemonics, which we hope will not be relevant for you, include "Oh Brother, Astronomers Frequently Give Killer Midterms" and "Oh Boy, An F Grade Kills Me!" With the new L, T, and Y spectral classes, the mnemonic might be expanded to "Oh Be A Fine Girl (Guy), Kiss Me Like That, Yo!"

Each of these spectral classes, except possibly for the Y class which is still being defined, is further subdivided into 10 subclasses designated by the numbers 0 through 9. A B0 star is the hottest type of B star; a B9 star is the coolest type of B star and is only slightly hotter than an A0 star.

And just one more item of vocabulary: for historical reasons, astronomers call all the elements heavier than helium *metals*, even though most of them do not show metallic properties. (If you are getting annoyed at the peculiar jargon that astronomers use, just bear in mind that every field of human activity tends to develop its own specialized vocabulary. Just try reading a credit card or social media agreement form these days without training in law!)

Let’s take a look at some of the details of how the spectra of the stars change with temperature. (It is these details that allowed Annie Cannon to identify the spectral types of stars as quickly as three per minute!) As [link] shows, in the hottest O stars (those with temperatures over 28,000 K), only lines of ionized helium and highly ionized atoms of other elements are conspicuous. Hydrogen lines are strongest in A stars with atmospheric temperatures of about 10,000 K. Ionized metals provide the most conspicuous lines in stars with temperatures from 6000 to 7500 K (spectral type F). In the coolest M stars (below 3500 K), absorption bands of titanium oxide and other molecules are very strong. By the way, the spectral class assigned to the Sun is G2. The sequence of spectral classes is summarized in [link].

Absorption Lines in Stars of Different Temperatures.



This graph shows the strengths of absorption lines of different chemical species (atoms, ions, molecules) as we move from hot (left) to cool (right) stars. The sequence of spectral types is also shown.

Spectral Classes for Stars				
Spectral Class	**Color**	**Approximate Temperature (K)**	**Principal Features**	**Examples**
O	Blue	> 30,000	Neutral and ionized helium lines, weak hydrogen lines	10 Lacertae
B	Blue-white	10,000–30,000	Neutral helium lines, strong hydrogen lines	Rigel, Spica
A	White	7500–10,000	Strongest hydrogen lines, weak ionized calcium lines, weak ionized metal (e.g., iron, magnesium) lines	Sirius, Vega

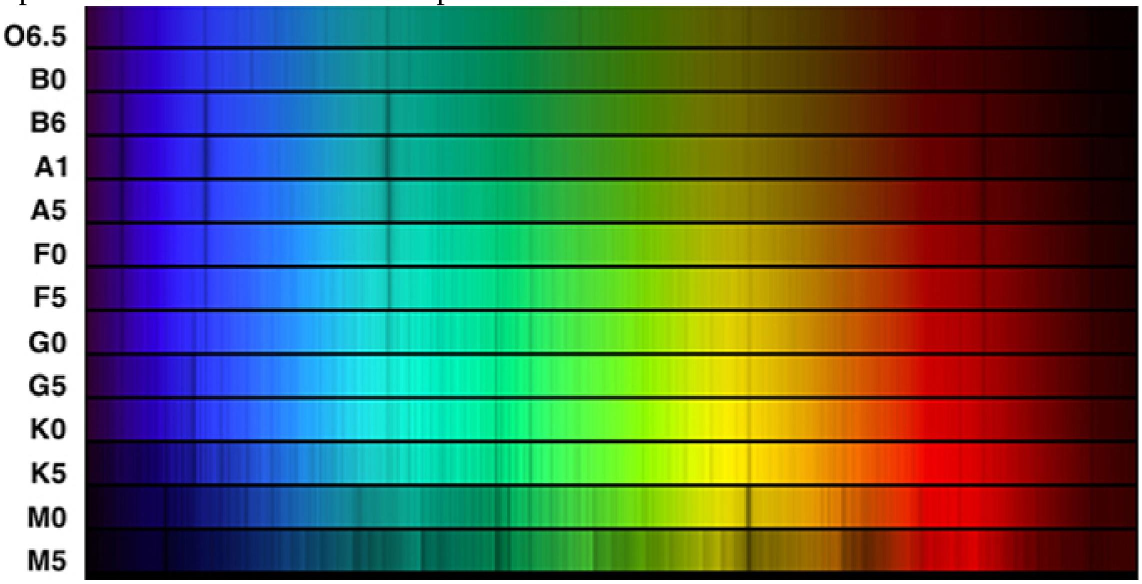
Spectral Classes for Stars				
Spectral Class	**Color**	**Approximate Temperature (K)**	**Principal Features**	**Examples**
F	Yellow-white	6000–7500	Strong hydrogen lines, strong ionized calcium lines, weak sodium lines, many ionized metal lines	Canopus, Procyon
G	Yellow	5200–6000	Weaker hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of ionized and neutral metals	Sun, Capella
K	Orange	3700–5200	Very weak hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of neutral metals	Arcturus, Aldebaran

Spectral Classes for Stars				
Spectral Class	**Color**	**Approximate Temperature (K)**	**Principal Features**	**Examples**
M	Red	2400–3700	Strong lines of neutral metals and molecular bands of titanium oxide dominate	Betelgeuse, Antares
L	Red	1300–2400	Metal hydride lines, alkali metal lines (e.g., sodium, potassium, rubidium)	Teide 1
T	Magenta	700–1300	Methane lines	Gliese 229B
Y	Infrared[footnote] Absorption by sodium and potassium atoms makes Y dwarfs appear a bit less red than L dwarfs.	< 700	Ammonia lines	WISE 1828+2650

To see how spectral classification works, let’s use [link]. Suppose you have a spectrum in which the hydrogen lines are about half as strong as those seen in an A star. Looking at the lines in our figure, you see that the star could be either a B star or a G star. But if the spectrum also contains helium lines, then it is a B star, whereas if it contains lines of ionized iron and other metals, it must be a G star.

If you look at [link], you can see that you, too, could assign a spectral class to a star whose type was not already known. All you have to do is match the pattern of spectral lines to a

standard star (like the ones shown in the figure) whose type has already been determined.
Spectra of Stars with Different Spectral Classes.



This image compares the spectra of the different spectral classes. The spectral class assigned to each of these stellar spectra is listed at the left of the picture. The strongest four lines seen at spectral type A1 (one in the red, one in the blue-green, and two in the blue) are Balmer lines of hydrogen. Note how these lines weaken at both higher and lower temperatures, as [link] also indicates. The strong pair of closely spaced lines in the yellow in the cool stars is due to neutral sodium (one of the neutral metals in [link]). (Credit: modification of work by NOAO/AURA/NSF)

Both colors and spectral classes can be used to estimate the temperature of a star. Spectra are harder to measure because the light has to be bright enough to be spread out into all colors of the rainbow, and detectors must be sensitive enough to respond to individual wavelengths. In order to measure colors, the detectors need only respond to the many wavelengths that pass simultaneously through the colored filters that have been chosen—that is, to *all* the blue light or *all* the yellow-green light.

Note:
Annie Cannon: Classifier of the Stars
Annie Jump Cannon was born in Delaware in 1863 ([link]). In 1880, she went to Wellesley College, one of the new breed of US colleges opening up to educate young women. Wellesley, only 5 years old at the time, had the second student physics lab in the country and provided excellent training in basic science. After college, Cannon spent a decade with her parents but was very dissatisfied, longing to do scientific work. After her

mother’s death in 1893, she returned to Wellesley as a teaching assistant and also to take courses at Radcliffe, the women’s college associated with Harvard.

Annie Jump Cannon (1863–1941).

Cannon is well-known for her classifications of stellar spectra. (credit: modification of work by Smithsonian Institution)

In the late 1800s, the director of the Harvard Observatory, Edward C. Pickering, needed lots of help with his ambitious program of classifying stellar spectra. The basis for these studies was a monumental collection of nearly a million photographic spectra of stars, obtained from many years of observations made at Harvard College Observatory in Massachusetts as well as at its remote observing stations in South America and South Africa. Pickering quickly discovered that educated young women could be hired as assistants for one-third or one-fourth the salary paid to men, and they would often put up with working conditions and repetitive tasks that men with the same education would not tolerate. These women became known as the Harvard Computers. (We should emphasize that astronomers were not alone in reaching such conclusions about the relatively new idea of upper-class, educated women working outside the home: women were exploited and undervalued in many fields. This is a legacy from which our society is just beginning to emerge.)

Cannon was hired by Pickering as one of the “computers” to help with the classification of spectra. She became so good at it that she could visually examine and determine the spectral types of several hundred stars per hour (dictating her conclusions to an assistant). She made many discoveries while investigating the Harvard photographic plates, including 300 variable stars (stars whose luminosity changes periodically). But her main legacy is a marvelous catalog of spectral types for hundreds of thousands of stars, which served as a foundation for much of twentieth-century astronomy.

In 1911, a visiting committee of astronomers reported that “she is the one person in the world who can do this work quickly and accurately” and urged Harvard to give Cannon an official appointment in keeping with her skill and renown. Not until 1938, however, did Harvard appoint her an astronomer at the university; she was then 75 years old.

Cannon received the first honorary degree Oxford awarded to a woman, and she became the first woman to be elected an officer of the American Astronomical Society, the main

professional organization of astronomers in the US. She generously donated the money from one of the major prizes she had won to found a special award for women in astronomy, now known as the Annie Jump Cannon Prize. True to form, she continued classifying stellar spectra almost to the very end of her life in 1941.

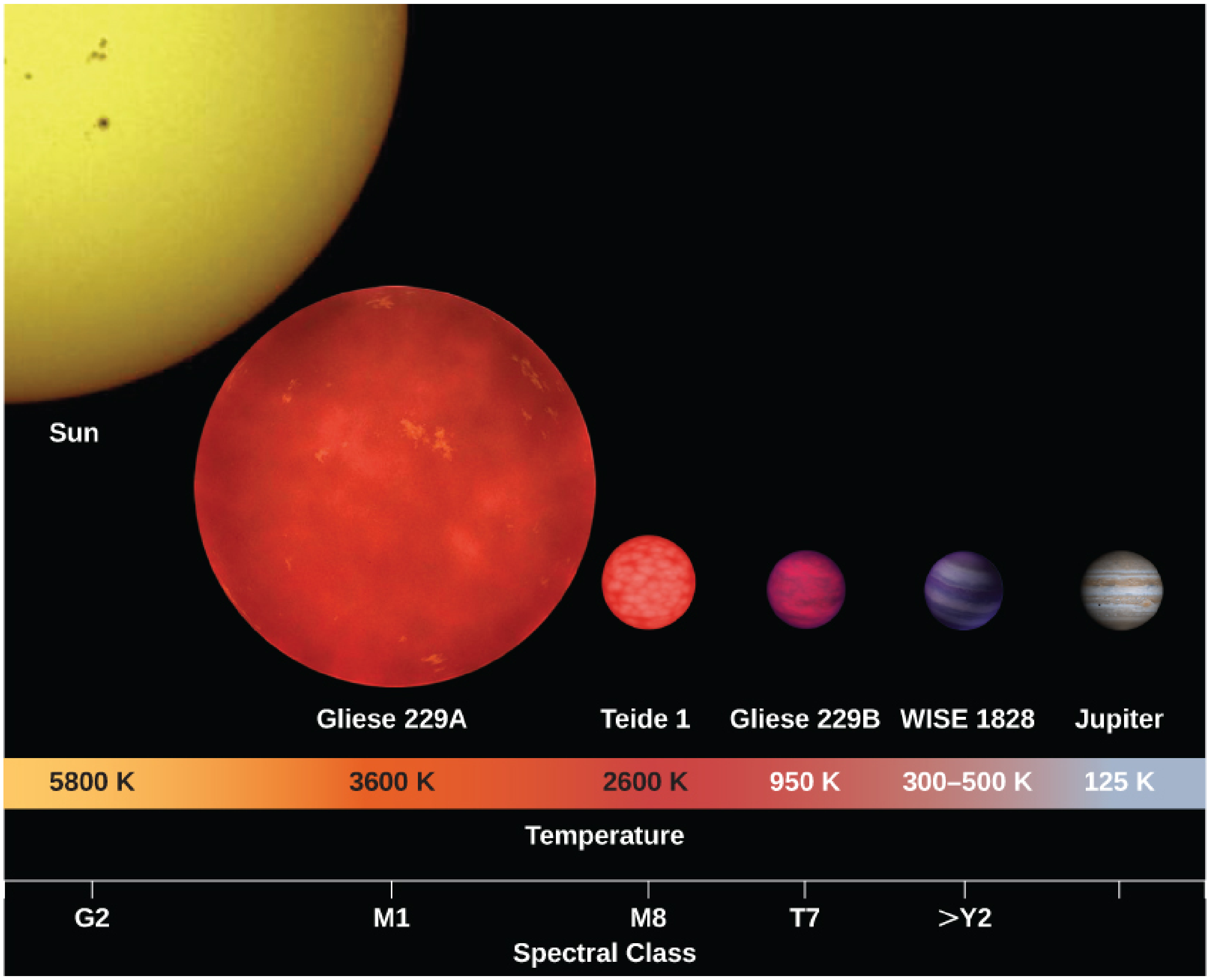
Spectral Classes L, T, and Y

The scheme devised by Cannon worked well until 1988, when astronomers began to discover objects even cooler than M9-type stars. We use the word *object* because many of the new discoveries are not true stars. A star is defined as an object that during some part of its lifetime derives 100% of its energy from the same process that makes the Sun shine—the fusion of hydrogen nuclei (protons) into helium. Objects with masses less than about 7.5% of the mass of our Sun (about 0.075 M_{Sun}) do not become hot enough for hydrogen fusion to take place. Even before the first such “failed star” was found, this class of objects, with masses intermediate between stars and planets, was given the name **brown dwarfs**.

Brown dwarfs are very difficult to observe because they are extremely faint and cool, and they put out most of their light in the infrared part of the spectrum. It was only after the construction of very large telescopes, like the Keck telescopes in Hawaii, and the development of very sensitive infrared detectors, that the search for brown dwarfs succeeded. The first brown dwarf was discovered in 1988, and, as of the summer of 2015, there are more than 2200 known brown dwarfs.

Initially, brown dwarfs were given spectral classes like $M10^+$ or “much cooler than M9,” but so many are now known that it is possible to begin assigning spectral types. The hottest brown dwarfs are given types L0–L9 (temperatures in the range 2400–1300 K), whereas still cooler (1300–700 K) objects are given types T0–T9 (see [link]). In class L brown dwarfs, the lines of titanium oxide, which are strong in M stars, have disappeared. This is because the L dwarfs are so cool that atoms and molecules can gather together into dust particles in their atmospheres; the titanium is locked up in the dust grains rather than being available to form molecules of titanium oxide. Lines of steam (hot water vapor) are present, along with lines of carbon monoxide and neutral sodium, potassium, cesium, and rubidium. Methane (CH_4) lines are strong in class-T brown dwarfs, as methane exists in the atmosphere of the giant planets in our own solar system.

In 2009, astronomers discovered ultra-cool brown dwarfs with temperatures of 500–600 K. These objects exhibited absorption lines due to ammonia (NH_3), which are not seen in T dwarfs. A new spectral class, Y, was created for these objects. As of 2015, over two dozen brown dwarfs belonging to spectral class Y have been discovered, some with temperatures comparable to that of the human body (about 300 K).
Brown Dwarfs.



This illustration shows the sizes and surface temperatures of brown dwarfs Teide 1, Gliese 229B, and WISE1828 in relation to the Sun, a red dwarf star (Gliese 229A), and Jupiter. (credit: modification of work by MPIA/V. Joergens)

Most brown dwarfs start out with atmospheric temperatures and spectra like those of true stars with spectral classes of M6.5 and later, even though the brown dwarfs are not hot and dense enough in their interiors to fuse hydrogen. In fact, the spectra of brown dwarfs and true stars are so similar from spectral types late M through L that it is not possible to distinguish the two types of objects based on spectra alone. An independent measure of mass is required to determine whether a specific object is a brown dwarf or a very low mass star. Since brown dwarfs cool steadily throughout their lifetimes, the spectral type of a given brown dwarf changes with time over a billion years or more from late M through L, T, and Y spectral types.

Low-Mass Brown Dwarfs vs. High-Mass Planets

An interesting property of brown dwarfs is that they are all about the same radius as Jupiter, regardless of their masses. Amazingly, this covers a range of masses from about 13 to 80 times the mass of Jupiter (M_J). This can make distinguishing a low-mass brown dwarf from a high-mass planet very difficult.

So, what is the difference between a low-mass brown dwarf and a high-mass planet? The International Astronomical Union considers the distinctive feature to be *deuterium fusion*. Although brown dwarfs do not sustain regular (proton-proton) hydrogen fusion, they are capable of fusing deuterium (a rare form of hydrogen with one proton and one neutron in its nucleus). The fusion of deuterium can happen at a lower temperature than the fusion of hydrogen. If an object has enough mass to fuse deuterium (about 13 M_J or 0.012 M_{Sun}), it is a brown dwarf. Objects with less than 13 M_J do not fuse deuterium and are usually considered planets.

Key Concepts and Summary

The differences in the spectra of stars are principally due to differences in temperature, not composition. The spectra of stars are described in terms of spectral classes. In order of decreasing temperature, these spectral classes are O, B, A, F, G, K, M, L, T, and Y. These are further divided into subclasses numbered from 0 to 9. The classes L, T, and Y have been added recently to describe newly discovered star-like objects—mainly brown dwarfs—that are cooler than M9. Our Sun has spectral type G2.

Glossary

brown dwarf
: an object intermediate in size between a planet and a star; the approximate mass range is from about 1/100 of the mass of the Sun up to the lower mass limit for self-sustaining nuclear reactions, which is about 0.075 the mass of the Sun; brown dwarfs are capable of deuterium fusion, but not hydrogen fusion

spectral class
: (or spectral type) the classification of stars according to their temperatures using the characteristics of their spectra; the types are O, B, A, F, G, K, and M with L, T, and Y added recently for cooler star-like objects that recent survey have revealed

Using Spectra to Measure Stellar Radius, Composition, and Motion

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Understand how astronomers can learn about a star's radius and composition by studying its spectrum
- Explain how astronomers can measure the motion and rotation of a star using the Doppler effect
- Describe the proper motion of a star and how it relates to a star's space velocity

Analyzing the spectrum of a star can teach us all kinds of things in addition to its temperature. We can measure its detailed chemical composition as well as the pressure in its atmosphere. From the pressure, we get clues about its size. We can also measure its motion toward or away from us and estimate its rotation.

Clues to the Size of a Star

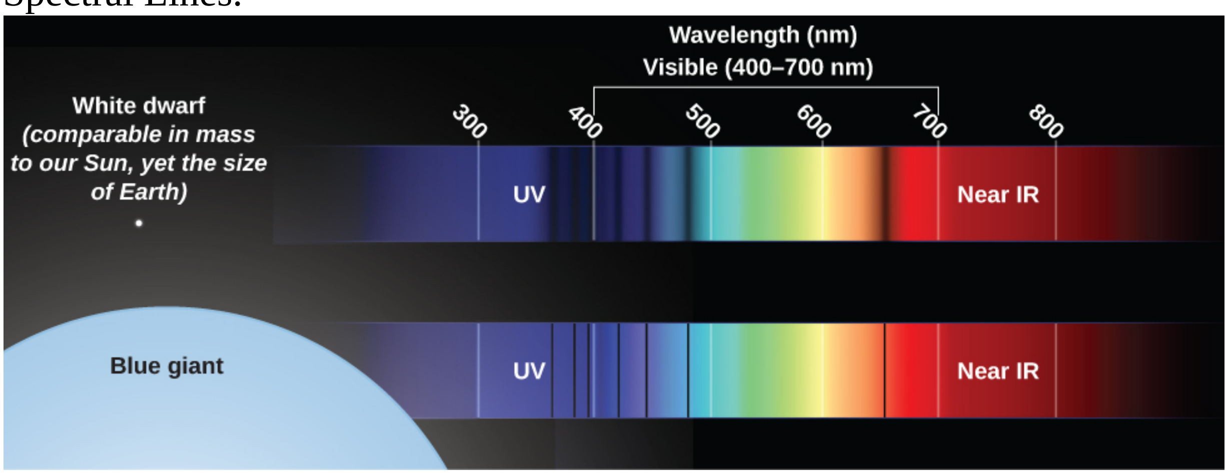
As we shall see in The Stars: A Celestial Census, stars come in a wide variety of sizes. At some periods in their lives, stars can expand to enormous dimensions. Stars of such exaggerated size are called **giants**. Luckily for the astronomer, stellar spectra can be used to distinguish giants from run-of-the-mill stars (such as our Sun).

Suppose you want to determine whether a star is a giant. A giant star has a large, extended photosphere. Because it is so large, a giant star's atoms are

spread over a great volume, which means that the density of particles in the star's photosphere is low. As a result, the pressure in a giant star's photosphere is also low. This low pressure affects the spectrum in two ways. First, a star with a lower-pressure photosphere shows narrower spectral lines than a star of the same temperature with a higher-pressure photosphere ([link]). The difference is large enough that careful study of spectra can tell which of two stars at the same temperature has a higher pressure (and is thus more compressed) and which has a lower pressure (and thus must be extended). This effect is due to collisions between particles in the star's photosphere—more collisions lead to broader spectral lines. Collisions will, of course, be more frequent in a higher-density environment. Think about it like traffic—collisions are much more likely during rush hour, when the density of cars is high.

Second, more atoms are ionized in a giant star than in a star like the Sun with the same temperature. The ionization of atoms in a star's outer layers is caused mainly by photons, and the amount of energy carried by photons is determined by temperature. But how long atoms *stay* ionized depends in part on pressure. Compared with what happens in the Sun (with its relatively dense photosphere), ionized atoms in a giant star's photosphere are less likely to pass close enough to electrons to interact and combine with one or more of them, thereby becoming neutral again. Ionized atoms, as we discussed earlier, have different spectra from atoms that are neutral.

Spectral Lines.



This figure illustrates one difference in the spectral lines from stars of the same temperature but different pressures. A giant star with a very-

low-pressure photosphere shows very narrow spectral lines (bottom), whereas a smaller star with a higher-pressure photosphere shows much broader spectral lines (top). (credit: modification of work by NASA, ESA, A. Field, and J. Kalirai (STScI))

Abundances of the Elements

Absorption lines of a majority of the known chemical elements have now been identified in the spectra of the Sun and stars. If we see lines of iron in a star's spectrum, for example, then we know immediately that the star must contain iron.

Note that the *absence* of an element's spectral lines does not necessarily mean that the element itself is absent. As we saw, the temperature and pressure in a star's atmosphere will determine what types of atoms are able to produce absorption lines. Only if the physical conditions in a star's photosphere are such that lines of an element *should* (according to calculations) be there can we conclude that the absence of observable spectral lines implies low abundance of the element.

Suppose two stars have identical temperatures and pressures, but the lines of, say, sodium are stronger in one than in the other. Stronger lines mean that there are more atoms in the stellar photosphere absorbing light. Therefore, we know immediately that the star with stronger sodium lines contains more sodium. Complex calculations are required to determine exactly how much more, but those calculations can be done for any element observed in any star with any temperature and pressure.

Of course, astronomy textbooks such as ours always make these things sound a bit easier than they really are. If you look at the stellar spectra such as those in [link], you may get some feeling for how hard it is to decode all of the information contained in the thousands of absorption lines. First of all, it has taken many years of careful laboratory work on Earth to determine the precise wavelengths at which hot gases of each element have their spectral lines. Long books and computer databases have been

compiled to show the lines of each element that can be seen at each temperature. Second, stellar spectra usually have many lines from a number of elements, and we must be careful to sort them out correctly. Sometimes nature is unhelpful, and lines of different elements have identical wavelengths, thereby adding to the confusion. And third, as we saw in the chapter on Radiation and Spectra, the motion of the star can change the observed wavelength of each of the lines. So, the observed wavelengths may not match laboratory measurements exactly. In practice, analyzing stellar spectra is a demanding, sometimes frustrating task that requires both training and skill.

Studies of stellar spectra have shown that hydrogen makes up about three-quarters of the mass of most stars. Helium is the second-most abundant element, making up almost a quarter of a star's mass. Together, hydrogen and helium make up from 96 to 99% of the mass; in some stars, they amount to more than 99.9%. Among the 4% or less of "heavy elements," oxygen, carbon, neon, iron, nitrogen, silicon, magnesium, and sulfur are among the most abundant. Generally, but not invariably, the elements of lower atomic weight are more abundant than those of higher atomic weight.

Take a careful look at the list of elements in the preceding paragraph. Two of the most abundant are hydrogen and oxygen (which make up water); add carbon and nitrogen and you are starting to write the prescription for the chemistry of an astronomy student. We are made of elements that are common in the universe—just mixed together in a far more sophisticated form (and a much cooler environment) than in a star.

As we mentioned in The Spectra of Stars (and Brown Dwarfs) section, astronomers use the term "metals" to refer to all elements heavier than hydrogen and helium. The fraction of a star's mass that is composed of these elements is referred to as the star's *metallicity*. The metallicity of the Sun, for example, is 0.02, since 2% of the Sun's mass is made of elements heavier than helium.

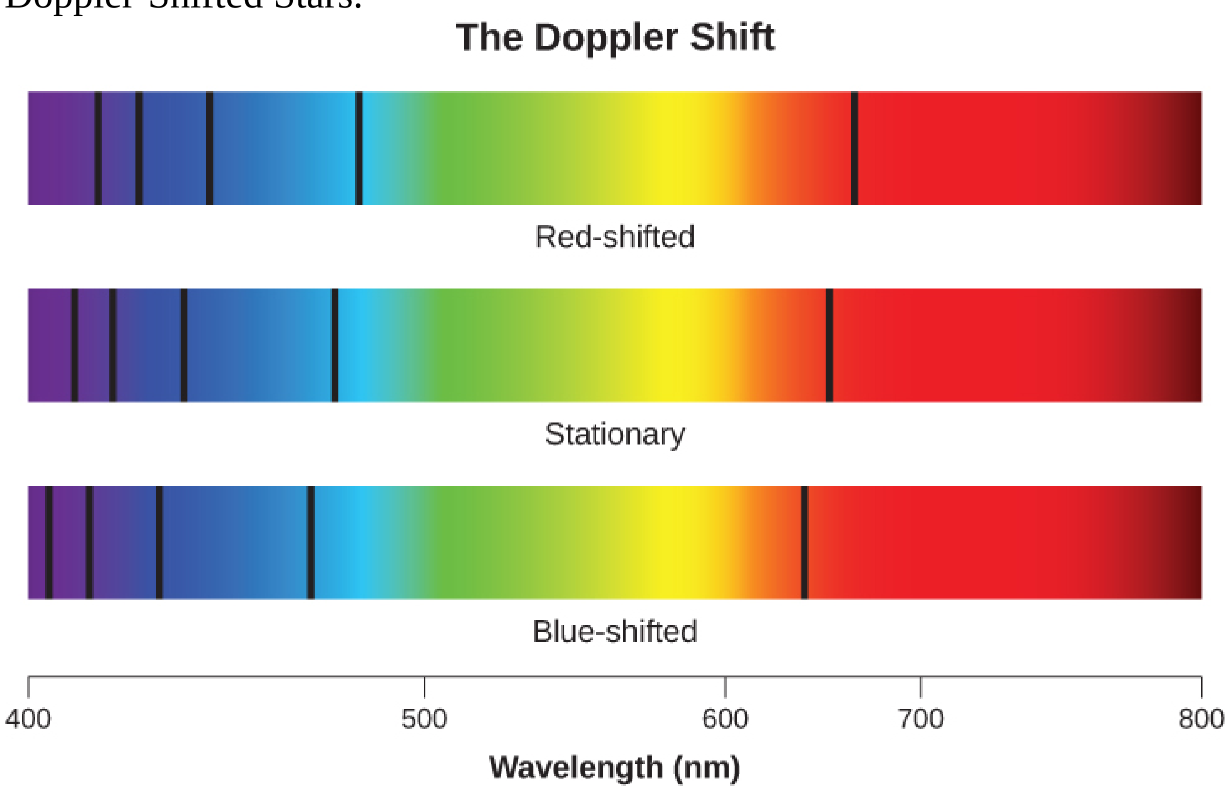
Appendix K lists how common each element is in the universe (compared to hydrogen); these estimates are based primarily on investigation of the Sun, which is a typical star. Some very rare elements, however, have not been detected in the Sun. Estimates of the amounts of these elements in the

universe are based on laboratory measurements of their abundance in primitive meteorites, which are considered representative of unaltered material condensed from the solar nebula (see the Cosmic Samples and the Origin of the Solar System chapter).

Radial Velocity

When we measure the spectrum of a star, we determine the wavelength of each of its lines. If the star is not moving with respect to the Sun, then the wavelength corresponding to each element will be the same as those we measure in a laboratory here on Earth. But if stars are moving toward or away from us, we must consider the *Doppler effect* (see The **Doppler** Effect section). We should see all the spectral lines of moving stars shifted toward the red end of the spectrum if the star is moving away from us, or toward the blue (violet) end if it is moving toward us ([link]). The greater the shift, the faster the star is moving. Such motion, along the line of sight between the star and the observer, is called **radial velocity** and is usually measured in kilometers per second.

Doppler-Shifted Stars.



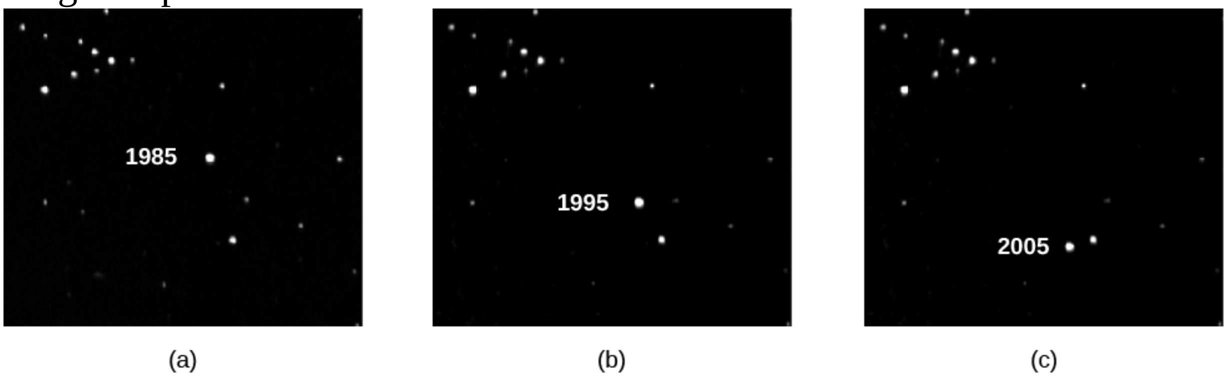
When the spectral lines of a moving star shift toward the red end of the spectrum, we know that the star is moving away from us. If they shift toward the blue end, the star is moving toward us.

William Huggins, pioneering yet again, in 1868 made the first radial velocity determination of a star. He observed the Doppler shift in one of the hydrogen lines in the spectrum of Sirius and found that this star is moving toward the solar system. (Better measurements now indicate that Sirius is actually moving away, but Huggins is still celebrated as a pioneer in making such measurements.) Today, radial velocity can be measured for any star bright enough for its spectrum to be observed. As we will see in The Stars: A Celestial Census, radial velocity measurements of double stars are crucial in deriving stellar masses.

Proper Motion

There is another type of motion stars can have that cannot be detected with stellar spectra. Unlike radial motion, which is along our line of sight (i.e., toward or away from Earth), this motion, called **proper motion**, is *transverse*: that is, across our line of sight. We see it as a change in the relative positions of the stars on the celestial sphere ([link]). These changes are very slow. Even the star with the largest proper motion takes 200 years to change its position in the sky by an amount equal to the width of the full Moon, and the motions of other stars are smaller yet.

Large Proper Motion.

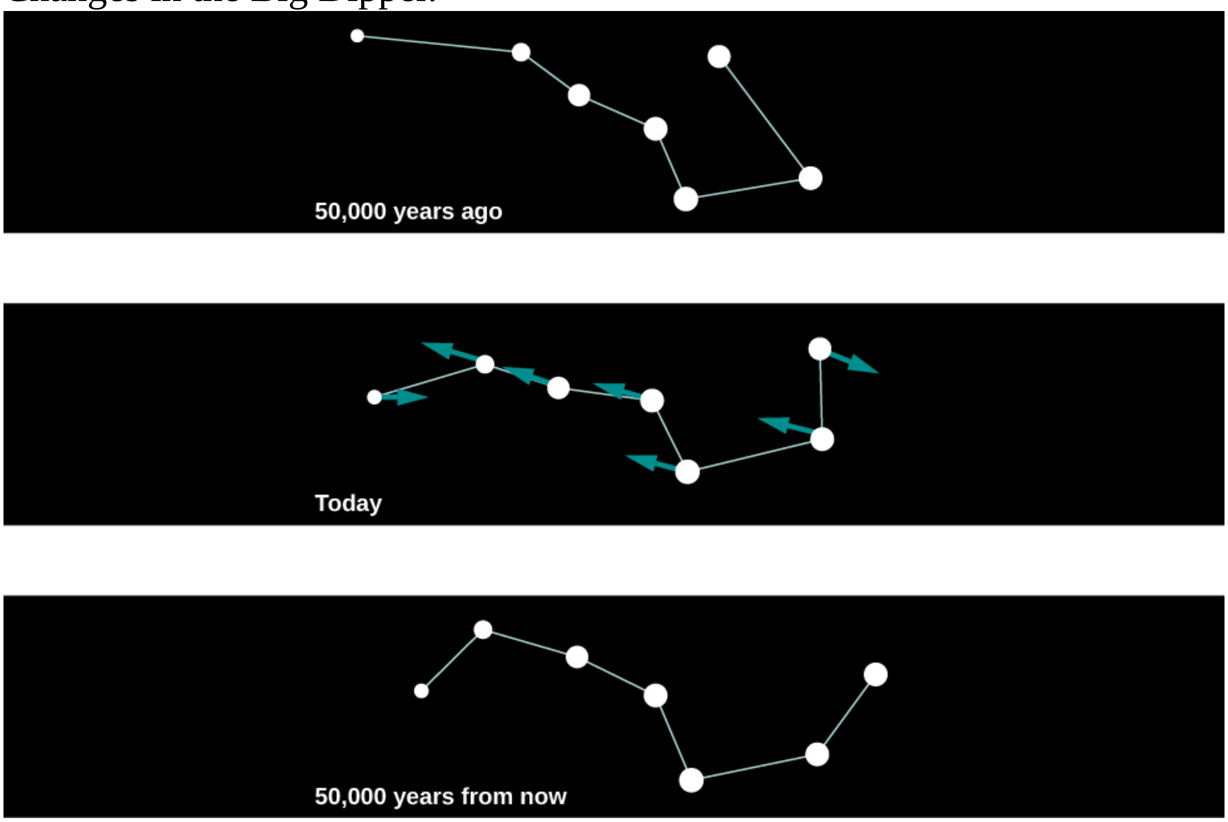


Three photographs of Barnard’s star, the star with the largest known

proper motion, show how this faint star has moved over a period of 20 years. (modification of work by Steve Quirk)

For this reason, with our naked eyes, we do not notice any change in the positions of the bright stars during the course of a human lifetime. If we could live long enough, however, the changes would become obvious. For example, some 50,000 years from now, terrestrial observers will find the handle of the Big Dipper unmistakably more bent than it is now ([link]).

Changes in the Big Dipper.



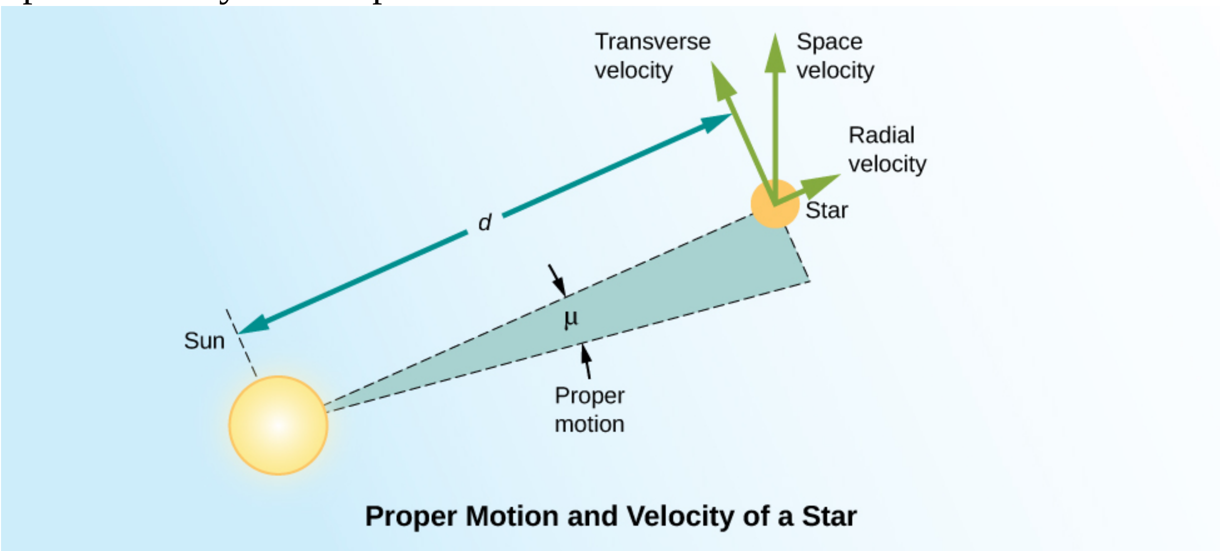
This figure shows changes in the appearance of the Big Dipper due to proper motion of the stars over 100,000 years.

We measure the proper motion of a star in arcseconds (1/3600 of a degree) per year. That is, the measurement of proper motion tells us only by how

much of an angle a star has changed its position on the celestial sphere. If two stars at different distances are moving at the same velocity perpendicular to our line of sight, the closer one will show a larger shift in its position on the celestial sphere in a year's time. As an analogy, imagine you are standing at the side of a freeway. Cars will appear to whiz past you. If you then watch the traffic from a vantage point half a mile away, the cars will move much more slowly across your field of vision. In order to convert this angular motion to a velocity, we need to know how far away the star is.

To know the true **space velocity** of a star—that is, its total speed and the direction in which it is moving through space relative to the Sun—we must know its radial velocity, proper motion, and distance ([link]). A star's space velocity can also, over time, cause its distance from the Sun to change significantly. Over several hundred thousand years, these changes can be large enough to affect the apparent brightnesses of nearby stars. Today, Sirius, in the constellation Canis Major (the Big Dog) is the brightest star in the sky, but 100,000 years ago, the star Canopus in the constellation Carina (the Keel) was the brightest one. A little over 200,000 years from now, Sirius will have moved away and faded somewhat, and Vega, the bright blue star in Lyra, will take over its place of honor as the brightest star in Earth's skies.

Space Velocity and Proper Motion.



This figure shows the true space velocity of a star. The radial velocity is the component of the space velocity projected along the line of sight

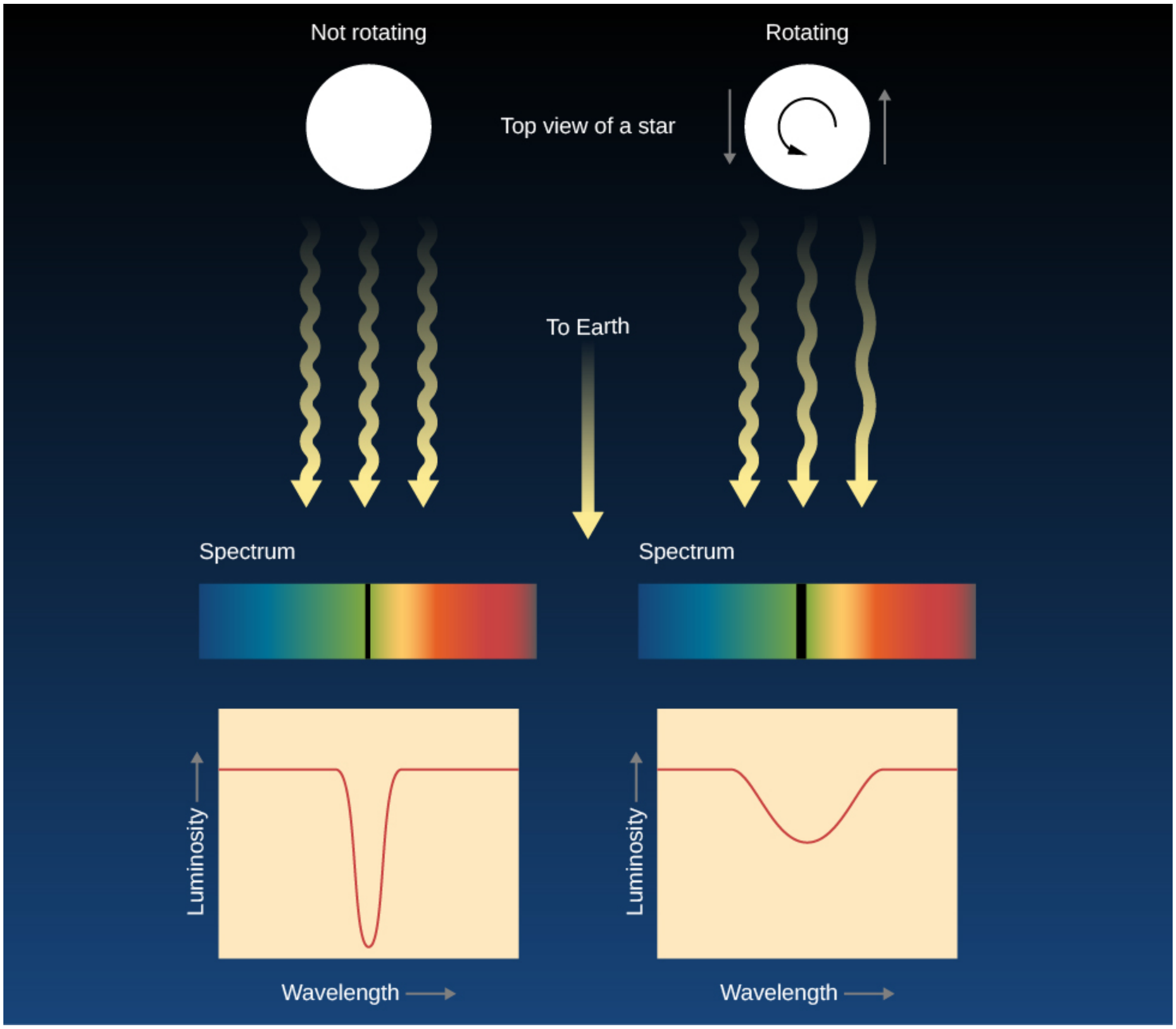
from the Sun to a star. The transverse velocity is a component of the space velocity projected on the sky. What astronomers measure is proper motion (μ), which is the change in the apparent direction on the sky measured in fractions of a degree. To convert this change in direction to a speed in, say, kilometers per second, it is necessary to also know the distance (d) from the Sun to the star.

Rotation

We can also use the Doppler effect to measure how fast a star rotates. If an object is rotating, then one of its sides is approaching us while the other is receding (unless its axis of rotation happens to be pointed exactly toward us). This is clearly the case for the Sun or a planet; we can observe the light from either the approaching or receding edge of these nearby objects and directly measure the Doppler shifts that arise from the rotation.

Stars, however, are so far away that they all appear as unresolved points. The best we can do is to analyze the light from the entire star at once. Due to the Doppler effect, the lines in the light that come from the side of the star rotating toward us are shifted to shorter wavelengths and the lines in the light from the opposite edge of the star are shifted to longer wavelengths. You can think of each spectral line that we observe as the sum or composite of spectral lines originating from different speeds with respect to us. Each point on the star has its own Doppler shift, so the absorption line we see from the whole star is actually much wider than it would be if the star were not rotating. If a star is rotating rapidly, there will be a greater spread of Doppler shifts and all its spectral lines should be quite broad. In fact, astronomers call this effect *line broadening*, and the amount of broadening can tell us the speed at which the star rotates ([link]).

Using a Spectrum to Determine Stellar Rotation.

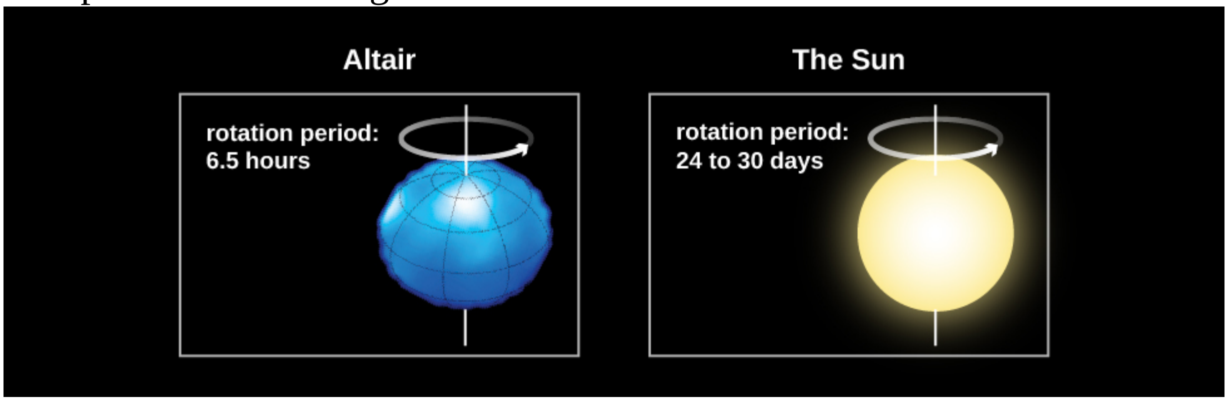


A rotating star will show broader spectral lines than a nonrotating star.

Measurements of the widths of spectral lines show that many stars rotate faster than the Sun, some with periods of less than a day! These rapid rotators spin so fast that their shapes are "flattened" into what we call *oblate spheroids*. An example of this is the star Vega, which rotates once every 12.5 hours. Vega's rotation flattens its shape so much that its diameter at the equator is 23% wider than its diameter at the poles ([link]). The Sun, with its rotation period of about a month, rotates rather slowly. Studies have shown that stars decrease their rotational speed as they age. Young stars

rotate very quickly, with rotational periods of days or less. Very old stars can have rotation periods of several months.

Comparison of Rotating Stars.



This illustration compares the more rapidly rotating star Altair to the slower rotating Sun.

As you can see, spectroscopy is an extremely powerful technique that helps us learn all kinds of information about stars that we simply could not gather any other way. We will see in later chapters that these same techniques can also teach us about galaxies, which are the most distant objects that can we observe. Without spectroscopy, we would know next to nothing about the universe beyond the solar system.

Note:

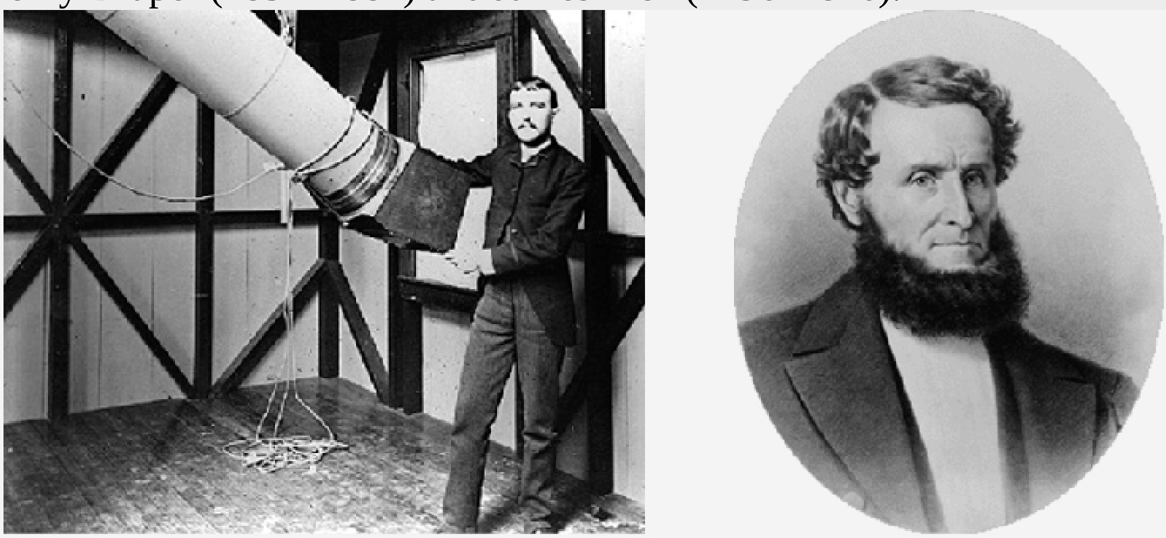
Astronomy and Philanthropy

Throughout the history of astronomy, contributions from wealthy patrons of the science have made an enormous difference in building new instruments and carrying out long-term research projects. Edward Pickering’s stellar classification project, which was to stretch over several decades, was made possible by major donations from Anna Draper. She was the widow of Henry Draper, a physician who was one of the most accomplished amateur astronomers of the nineteenth century and the first person to successfully photograph the spectrum of a star. Anna Draper gave several hundred thousand dollars to Harvard Observatory. As a result,

the great spectroscopic survey is still known as the Henry Draper Memorial, and many stars are still referred to by their “HD” numbers in that catalog (such as HD 209458).

In the 1870s, the eccentric piano builder and real estate magnate James Lick ([link]) decided to leave some of his fortune to build the world’s largest telescope. When, in 1887, the pier to house the telescope was finished, Lick’s body was entombed in it. Atop the foundation rose a 36-inch refractor, which for many years was the main instrument at the Lick Observatory near San Jose.

Henry Draper (1837–1882) and James Lick (1796–1876).



(a) (b)

(a) Draper stands next to a telescope used for photography. After his death, his widow funded further astronomy work in his name. (b) Lick was a philanthropist who provided funds to build a 36-inch refractor not only as a memorial to himself but also to aid in further astronomical research.

The Lick telescope remained the largest in the world until 1897, when George Ellery Hale persuaded railroad millionaire Charles Yerkes to finance the construction of a 40-inch telescope near Chicago. More recently, Howard Keck, whose family made its fortune in the oil industry, gave $70 million from his family foundation to the California Institute of Technology to help build the world’s largest telescope atop the 14,000-foot

peak of Mauna Kea in Hawaii (see the chapter on Astronomical Instruments to learn more about these telescopes). The Keck Foundation was so pleased with what is now called the Keck telescope that they gave $74 million more to build Keck II, another 10-meter reflector on the same volcanic peak.

Now, if any of you become millionaires or billionaires, and astronomy has sparked your interest, do keep an astronomical instrument or project in mind as you plan your estate. But frankly, private philanthropy could not possibly support the full enterprise of scientific research in astronomy. Much of our exploration of the universe is financed by federal agencies such as the National Science Foundation and NASA in the United States, and by similar government agencies in the other countries. In this way, all of us, through a very small share of our tax dollars, are philanthropists for astronomy.

Key Concepts and Summary

Spectra of stars of the same temperature but different atmospheric pressures have subtle differences, so spectra can be used to determine whether a star has a large radius and low atmospheric pressure (a giant star) or a small radius and high atmospheric pressure. Stellar spectra can also be used to determine the chemical composition of stars; hydrogen and helium make up most of the mass of all stars. Measurements of line shifts produced by the Doppler effect indicate the radial velocity of a star. Broadening of spectral lines by the Doppler effect is a measure of rotational velocity. A star can also show proper motion, due to the component of a star’s space velocity across the line of sight.

For Further Exploration

Articles

Berman, B. “Magnitude Cum Laude.” *Astronomy* (December 1998): 92. How we measure the apparent brightnesses of stars is discussed.

Dvorak, J. “The Women Who Created Modern Astronomy [including Annie Cannon].” *Sky & Telescope* (August 2013): 28.

Hearnshaw, J. “Origins of the Stellar Magnitude Scale.” *Sky & Telescope* (November 1992): 494. A good history of how we have come to have this cumbersome system is discussed.

Hirshfeld, A. “The Absolute Magnitude of Stars.” *Sky & Telescope* (September 1994): 35.

Kaler, J. “Stars in the Cellar: Classes Lost and Found.” *Sky & Telescope* (September 2000): 39. An introduction is provided for spectral types and the new classes L and T.

Kaler, J. “Origins of the Spectral Sequence.” *Sky & Telescope* (February 1986): 129.

Skrutskie, M. “2MASS: Unveiling the Infrared Universe.” *Sky & Telescope* (July 2001): 34. This article focuses on an all-sky survey at 2 microns.

Sneden, C. “Reading the Colors of the Stars.” *Astronomy* (April 1989): 36. This article includes a discussion of what we learn from spectroscopy.

Steffey, P. “The Truth about Star Colors.” *Sky & Telescope* (September 1992): 266. The color index and how the eye and film “see” colors are discussed.

Tomkins, J. “Once and Future Celestial Kings.” *Sky & Telescope* (April 1989): 59. Calculating the motion of stars and determining which stars were, are, and will be brightest in the sky are discussed.

Websites

Discovery of Brown Dwarfs: http://w.astro.berkeley.edu/~basri/bdwarfs/SciAm-book.pdf.

Listing of Nearby Brown Dwarfs: http://www.solstation.com/stars/pc10bd.htm.

Spectral Types of Stars: http://www.skyandtelescope.com/astronomy-equipment/the-spectral-types-of-stars/.

Stellar Velocities https://www.e-education.psu.edu/astro801/content/l4_p7.html.

Unheard Voices! The Contributions of Women to Astronomy: A Resource Guide: http://multiverse.ssl.berkeley.edu/women and http://www.astrosociety.org/education/astronomy-resource-guides/women-in-astronomy-an-introductory-resource-guide/.

Videos

When You Are Just Too Small to be a Star: https://www.youtube.com/watch?v=zXCDsb4n4KU. 2013 Public Talk on Brown Dwarfs and Planets by Dr. Gibor Basri of the University of California–Berkeley (1:32:52).

Collaborative Group Activities

A. The Voyagers in Astronomy feature on Annie Cannon: Classifier of the Stars discusses some of the difficulties women who wanted to do astronomy faced in the first half of the twentieth century. What does your group think about the situation for women today? Do men and women have an equal chance to become scientists? Discuss with your group whether, in your experience, boys and girls were equally encouraged to do science and math where you went to school.
B. In the section on magnitudes in The Brightness of Stars, we discussed how this old system of classifying how bright different stars appear to the eye first developed. Your authors complained about the fact that this old system still has to be taught to every generation of new students. Can your group think of any other traditional systems of doing things in science and measurement where tradition rules even

though common sense says a better system could certainly be found. Explain. (Hint: Try Daylight Savings Time, or metric versus English units.)

C. Suppose you could observe a star that has only one spectral line. Could you tell what element that spectral line comes from? Make a list of reasons with your group about why you answered yes or no.
D. A wealthy alumnus of your college decides to give $50 million to the astronomy department to build a world-class observatory for learning more about the characteristics of stars. Have your group discuss what kind of equipment they would put in the observatory. Where should this observatory be located? Justify your answers. (You may want to refer back to the Astronomical Instruments chapter and to revisit this question as you learn more about the stars and equipment for observing them in future chapters.)
E. For some astronomers, introducing a new spectral type for the stars (like the types L, T, and Y discussed in the text) is similar to introducing a new area code for telephone calls. No one likes to disrupt the old system, but sometimes it is simply necessary. Have your group make a list of steps an astronomer would have to go through to persuade colleagues that a new spectral class is needed.

Review Questions

Exercise:

Problem:

What two factors determine how bright a star appears to be in the sky?

Exercise:

Problem: Explain why color is a measure of a star’s temperature.

Exercise:

Problem:

What is the main reason that the spectra of all stars are not identical? Explain.

Exercise:

Problem:

What elements are stars mostly made of? How do we know this?

Exercise:

Problem:

What did Annie Cannon contribute to the understanding of stellar spectra?

Exercise:

Problem:

Name five characteristics of a star that can be determined by measuring its spectrum. Explain how you would use a spectrum to determine these characteristics.

Exercise:

Problem:

How do objects of spectral types L, T, and Y differ from those of the other spectral types?

Exercise:

Problem:

Do stars that look brighter in the sky have larger or smaller magnitudes than fainter stars?

Exercise:

Problem:

The star Antares has an apparent magnitude of 1.0, whereas the star Procyon has an apparent magnitude of 0.4. Which star appears brighter in the sky?

Exercise:

Problem:

Based on their colors, which of the following stars is hottest? Which is coolest? Archenar (blue), Betelgeuse (red), Capella (yellow).

Exercise:

Problem: Order the seven basic spectral types from hottest to coldest.

Exercise:

Problem:

What is the defining difference between a brown dwarf and a true star?

Thought Questions

Exercise:

Problem:

If the star Sirius emits 23 times more energy than the Sun, why does the Sun appear brighter in the sky?

Exercise:

Problem:

How would two stars of equal luminosity—one blue and the other red—appear in an image taken through a filter that passes mainly blue light? How would their appearance change in an image taken through a filter that transmits mainly red light?

Exercise:

Problem:

[link] lists the temperature ranges that correspond to the different spectral types. What part of the star do these temperatures refer to? Why?

Exercise:

Problem:

Suppose you are given the task of measuring the colors of the brightest stars, listed in Appendix J, through three filters: the first transmits blue light, the second transmits yellow light, and the third transmits red light. If you observe the star Vega, it will appear equally bright through each of the three filters. Which stars will appear brighter through the blue filter than through the red filter? Which stars will appear brighter through the red filter? Which star is likely to have colors most nearly like those of Vega?

Exercise:

Problem:

Star X has lines of ionized helium in its spectrum, and star Y has bands of titanium oxide. Which is hotter? Why? The spectrum of star Z shows lines of ionized helium and also molecular bands of titanium oxide. What is strange about this spectrum? Can you suggest an explanation?

Exercise:

Problem:

The spectrum of the Sun has hundreds of strong lines of nonionized iron but only a few, very weak lines of helium. A star of spectral type B has very strong lines of helium but very weak iron lines. Do these differences mean that the Sun contains more iron and less helium than the B star? Explain.

Exercise:

Problem:

What are the approximate spectral classes of stars with the following characteristics?

A. Balmer lines of hydrogen are very strong; some lines of ionized metals are present.

B. The strongest lines are those of ionized helium.
C. Lines of ionized calcium are the strongest in the spectrum; hydrogen lines show only moderate strength; lines of neutral and metals are present.
D. The strongest lines are those of neutral metals and bands of titanium oxide.

Exercise:

Problem:

Look at the chemical elements in Appendix K. Can you identify any relationship between the abundance of an element and its atomic weight? Are there any obvious exceptions to this relationship?

Exercise:

Problem:

Appendix I lists some of the nearest stars. Are most of these stars hotter or cooler than the Sun? Do any of them emit more energy than the Sun? If so, which ones?

Exercise:

Problem:

Appendix J lists the stars that appear brightest in our sky. Are most of these hotter or cooler than the Sun? Can you suggest a reason for the difference between this answer and the answer to the previous question? (Hint: Look at the luminosities.) Is there any tendency for a correlation between temperature and luminosity? Are there exceptions to the correlation?

Exercise:

Problem:

What star appears the brightest in the sky (other than the Sun)? The second brightest? What color is Betelgeuse? Use Appendix J to find the answers.

Exercise:

Problem:

Suppose hominids one million years ago had left behind maps of the night sky. Would these maps represent accurately the sky that we see today? Why or why not?

Exercise:

Problem:

Why can only a lower limit to the rate of stellar rotation be determined from line broadening rather than the actual rotation rate? (Refer to [link].)

Exercise:

Problem:

Why do you think astronomers have suggested three different spectral types (L, T, and Y) for the brown dwarfs instead of M? Why was one not enough?

Exercise:

Problem:

Sam, a college student, just bought a new car. Sam’s friend Adam, a graduate student in astronomy, asks Sam for a ride. In the car, Adam remarks that the colors on the temperature control are wrong. Why did he say that?

(credit: modification of work by Michael Sheehan)

Exercise:

Problem:

Would a red star have a smaller or larger magnitude in a red filter than in a blue filter?

Exercise:

Problem:

Two stars have proper motions of one arcsecond per year. Star A is 20 light-years from Earth, and Star B is 10 light-years away from Earth. Which one has the faster velocity in space?

Exercise:

Problem:

Suppose there are three stars in space, each moving at 100 km/s. Star A is moving across (i.e., perpendicular to) our line of sight, Star B is moving directly away from Earth, and Star C is moving away from Earth, but at a 30° angle to the line of sight. From which star will you observe the greatest Doppler shift? From which star will you observe the smallest Doppler shift?

Exercise:

Problem:

What would you say to a friend who made this statement, "The visible-light spectrum of the Sun shows weak hydrogen lines and strong calcium lines. The Sun must therefore contain more calcium than hydrogen."?

Figuring for Yourself

Exercise:

Problem:

In Appendix J, how much more luminous is the most luminous of the stars than the least luminous?

For [link] through [link], use the equations relating magnitude and apparent brightness given in the section on the magnitude scale in The Brightness of Stars and [link].

Exercise:

Problem:

Verify that if two stars have a difference of five magnitudes, this corresponds to a factor of 100 in the ratio $\left(\frac{b_2}{b_1}\right)$; that 2.5 magnitudes corresponds to a factor of 10; and that 0.75 magnitudes corresponds to a factor of 2.

Exercise:

Problem:

As seen from Earth, the Sun has an apparent magnitude of about −26.7. What is the apparent magnitude of the Sun as seen from Saturn, about 10 AU away? (Remember that one AU is the distance from Earth to the Sun and that the brightness decreases as the inverse square of the distance.) Would the Sun still be the brightest star in the sky?

Exercise:

Problem:

An astronomer is investigating a faint star that has recently been discovered in very sensitive surveys of the sky. The star has a magnitude of 16. How much less bright is it than Antares, a star with magnitude roughly equal to 1?

Exercise:

Problem:

The center of a faint but active galaxy has magnitude 26. How much less bright does it look than the very faintest star that our eyes can see, roughly magnitude 6?

Exercise:

Problem:

You have enough information from this chapter to estimate the distance to Alpha Centauri, the second nearest star, which has an apparent magnitude of 0. Since it is a G2 star, like the Sun, assume it has the same luminosity as the Sun and the difference in magnitudes is a result only of the difference in distance. Estimate how far away Alpha Centauri is. Describe the necessary steps in words and then do the calculation. (As we will learn in the Celestial Distances chapter, this method—namely, assuming that stars with identical spectral types emit the same amount of energy—is actually used to estimate distances to stars.) If you assume the distance to the Sun is in AU, your answer will come out in AU.

Exercise:

Problem:

Do the previous problem again, this time using the information that the Sun is 150,000,000 km away. You will get a very large number of km as your answer. To get a better feeling for how the distances compare, try calculating the time it takes light at a speed of 299,338 km/s to travel from the Sun to Earth and from Alpha Centauri to Earth. For Alpha Centauri, figure out how long the trip will take in years as well as in seconds.

Exercise:

Problem:

Star A and Star B have different apparent brightnesses but identical luminosities. If Star A is 20 light-years away from Earth and Star B is 40 light-years away from Earth, which star appears brighter and by what factor?

Exercise:

Problem:

Star A and Star B have different apparent brightnesses but identical luminosities. Star A is 10 light-years away from Earth and appears 36 times brighter than Star B. How far away is Star B?

Exercise:

Problem:

The star Sirius A has an apparent magnitude of −1.5. Sirius A has a dim companion, Sirius B, which is 10,000 times less bright than Sirius A. What is the apparent magnitude of Sirius B? Can Sirius B be seen with the naked eye?

Exercise:

Problem:

Our Sun, a type G star, has a surface temperature of 5800 K. We know, therefore, that it is cooler than a type O star and hotter than a type M star. Given what you learned about the temperature ranges of these types of stars, how many times hotter than our Sun is the hottest type O star? How many times cooler than our Sun is the coolest type M star?

Glossary

giant
: a star of exaggerated size with a large, extended photosphere

proper motion
: the angular change per year in the direction of a star as seen from the Sun

radial velocity
: motion toward or away from the observer; the component of relative velocity that lies in the line of sight

space velocity

the total (three-dimensional) speed and direction with which an object is moving through space relative to the Sun

Thinking Ahead
class="introduction"
Globular Cluster M80.

This beautiful image shows a giant cluster of stars called Messier 80, located about 28,000 light-years from Earth. Such crowded groups, which astronomers call globular clusters, contain hundreds of thousands of stars, including some of the RR Lyrae variables discussed in this chapter. Especially obvious in this picture

are the bright red giants, which are stars similar to the Sun in mass that are nearing the ends of their lives. (credit: modification of work by The Hubble Heritage Team (AURA/ STScI/ NASA))

How large is the universe? What is the most distant object we can see? These are among the most fundamental questions astronomers can ask. But just as babies must crawl before they can take their first halting steps, so too

must we start with a more modest question: How far away are the stars? And even this question proves to be very hard to answer. After all, stars are mere points of light. Suppose you see a point of light in the darkness when you are driving on a country road late at night. How can you tell whether it is a nearby firefly, an oncoming motorcycle some distance away, or the porchlight of a house much farther down the road? It's not so easy, is it? Astronomers faced an even more difficult problem when they tried to estimate how far away the stars are.

In this chapter, we begin with the fundamental definitions of distances on Earth and then extend our reach outward to the stars. We will also examine the newest satellites that are surveying the night sky and discuss the special types of stars that can be used as trail markers to distant galaxies.

Fundamental Units of Distance

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Understand the importance of defining a standard distance unit
- Explain how the meter was originally defined and how it has changed over time
- Discuss how radar is used to measure distances to the other members of the solar system

The first measures of distances were based on human dimensions—the inch as the distance between knuckles on the finger, or the yard as the span from the extended index finger to the nose of the British king. Later, the requirements of commerce led to some standardization of such units, but each nation tended to set up its own definitions. It was not until the middle of the eighteenth century that any real efforts were made to establish a uniform, international set of standards.

The Metric System

One of the enduring legacies of the era of the French emperor Napoleon is the establishment of the *metric system* of units, officially adopted in France in 1799 and now used in most countries around the world. The fundamental metric unit of length is the *meter*, originally defined as one ten-millionth of the distance along Earth's surface from the equator to the pole. French astronomers of the seventeenth and eighteenth centuries were pioneers in determining the dimensions of Earth, so it was logical to use their information as the foundation of the new system.

Practical problems exist with a definition expressed in terms of the size of Earth, since anyone wishing to determine the distance from one place to another can hardly be expected to go out and re-measure the planet. Therefore, an intermediate standard meter consisting of a bar of platinum-iridium metal was set up in Paris. In 1889, by international agreement, this bar was defined to be exactly one meter in length, and

precise copies of the original meter bar were made to serve as standards for other nations.

Other units of length are derived from the meter. Thus, 1 kilometer (km) equals 1000 meters, 1 centimeter (cm) equals 1/100 meter, and so on. Even the old British and American units, such as the inch and the mile, are now defined in terms of the metric system.

Modern Redefinitions of the Meter

In 1960, the official definition of the meter was changed again. As a result of improved technology for generating spectral lines of precisely known wavelengths (see the chapter on Radiation and Spectra), the meter was redefined to equal 1,650,763.73 wavelengths of a particular atomic transition in the element krypton-86. The advantage of this redefinition is that anyone with a suitably equipped laboratory can reproduce a standard meter, without reference to any particular metal bar.

In 1983, the meter was defined once more, this time in terms of the velocity of light. Light in a vacuum can travel a distance of one meter in 1/299,792,458.6 second. Today, therefore, light travel time provides our basic unit of length. Put another way, a distance of *one light-second* (the amount of space light covers in one second) is defined to be 299,792,458.6 meters. That's almost 300 million meters that light covers in just one second; light really is *very* fast! We could just as well use the light-second as the fundamental unit of length, but for practical reasons (and to respect tradition), we have defined the meter as a small fraction of the light-second.

Distance within the Solar System

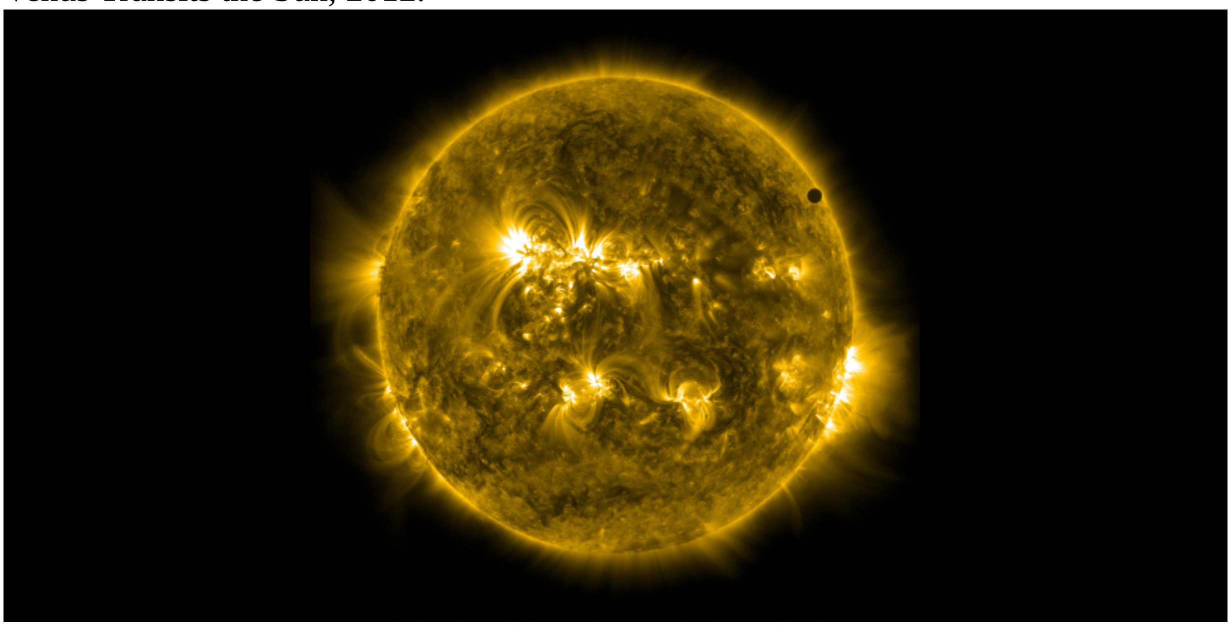
The work of Copernicus and Kepler established the *relative* distances of the planets—that is, how far from the Sun one planet is compared to another (see Observing the Sky: The Birth of Astronomy and Orbits and Gravity). But their work could not establish the *absolute* distances (in light-seconds or meters or other standard units of length). This is like knowing the height of all the students in your class only as compared to the height of your astronomy instructor, but not in inches or centimeters. Somebody's height has to be measured directly.

Similarly, to establish absolute distances, astronomers had to measure one distance in the solar system directly. Generally, the closer to us the object is, the easier such a measurement would be. Estimates of the distance to Venus were made as Venus crossed the face of the Sun in 1761 and 1769, and an international campaign was organized to estimate the distance to the asteroid Eros in the early 1930s, when its orbit brought it close to Earth. More recently, Venus crossed (or *transited*) the surface

of the Sun in 2004 and 2012, and allowed us to make a modern distance estimate, although, as we will see below, by then it wasn't needed ([link]).

Note:
If you would like more information on just how the motion of Venus across the Sun helped us pin down distances in the solar system, you can turn to a nice explanation by a NASA astronomer.

Venus Transits the Sun, 2012.



This striking "picture" of Venus crossing the face of the Sun (it's the black dot at about 2 o'clock) is more than just an impressive image. Taken with the Solar Dynamics Observatory spacecraft and special filters, it shows a modern transit of Venus. Such events allowed astronomers in the 1800s to estimate the distance to Venus. They measured the time it took Venus to cross the face of the Sun from different latitudes on Earth. The differences in times can be used to estimate the distance to the planet. Today, radar is used for much more precise distance estimates. (credit: modification of work by NASA/SDO, AIA)

The key to our modern determination of solar system dimensions is radar, a type of radio wave that can bounce off solid objects ([link]). As discussed in several earlier

chapters, by timing how long a radar beam (traveling at the speed of light) takes to reach another world and return, we can measure the distance involved very accurately. In 1961, radar signals were bounced off Venus for the first time, providing a direct measurement of the distance from Earth to Venus in terms of light-seconds (from the roundtrip travel time of the radar signal).

Subsequently, radar has been used to determine the distances to Mercury, Mars, the satellites of Jupiter, the rings of Saturn, and several asteroids. Note, by the way, that it is not possible to use radar to measure the distance to the Sun directly because the Sun does not reflect radar very efficiently. But we can measure the distance to many other solar system objects and use Kepler’s laws to give us the distance to the Sun.

Radar Telescope.

This dish-shaped antenna, part of the NASA Deep Space Network in California’s Mojave Desert, is 70 meters wide. Nicknamed the “Mars antenna,” this radar telescope can send and receive radar waves, and thus measure the distances to planets, satellites, and asteroids. (credit: NASA/JPL-Caltech)

From the various (related) solar system distances, astronomers selected the average distance from Earth to the Sun as our standard "measuring stick" within the solar system. When Earth and the Sun are closest, they are about 147.1 million kilometers apart; when Earth and the Sun are farthest, they are about 152.1 million kilometers apart. The average of these two distances is called the astronomical unit (AU). We then express all the other distances in the solar system in terms of the AU. Years of painstaking analyses of radar measurements have led to a determination of the length of the AU to a precision of about one part in a billion. The length of 1 AU can be expressed in light travel time as 499.004854 light-seconds, or about 8.3 light-minutes. If we use the definition of the meter given previously, this is equivalent to 1 AU = 149,597,870,700 meters.

These distances are, of course, given here to a much higher level of precision than is normally needed. In this text, we are usually content to express numbers to a couple of significant places and leave it at that. For our purposes, it will be sufficient to round off these numbers:

Equation:

We now know the absolute distance scale within our own solar system with fantastic accuracy. This is the first link in the chain of cosmic distances.

Note:

The distances between the celestial bodies in our solar system are sometimes difficult to grasp or put into perspective. This interactive website provides a "map" that shows the distances by using a scale at the bottom of the screen and allows you to scroll (using your arrow keys) through screens of "empty space" to get to the next planet—all while your current distance from the Sun is visible on the scale.

Key Concepts and Summary

Early measurements of length were based on human dimensions, but today, we use worldwide standards that specify lengths in units such as the meter. Distances within

the solar system are now determined by timing how long it takes radar signals to travel from Earth to the surface of a planet or other body and then return.

Thinking Ahead
class="introduction"
Where Stars Are Born.

We see a close-up of part of the Carina Nebula taken with the Hubble Space Telescope. This image reveals jets powered by newly forming stars embedded in a great cloud of gas and dust. Parts of the clouds are glowing from the energy of very young stars recently formed within them. (credit: modification of work by NASA,

ESA, and
M. Livio
and the
Hubble 20th
Anniversary
Team
(STScI))

"There are countless suns and countless earths all rotating round their suns in exactly the same way as the planets of our system. We see only the suns because they are the largest bodies and are luminous, but their planets remain invisible to us because they are smaller and non-luminous. . . . The unnumbered worlds in the universe are all similar in form and rank and subject to the same forces and the same laws."—Giordano Bruno in *On the Infinite Universe and Worlds* (1584). Bruno was tried for heresy by the Roman Inquisition and burned at the stake in 1600.

We've discussed stars as nuclear furnaces that convert light elements into heavier ones. A star's nuclear evolution begins when hydrogen is fused into helium, but that can only occur when the core temperature exceeds 10 to 12 million K. Since stars form from cold interstellar material, we must understand how they collapse and eventually reach this "ignition temperature" to explain the birth of stars. Star formation is a continuous process, from the birth of our Galaxy right up to today. We estimate that

every year in our Galaxy, on average, three solar masses of interstellar matter are converted into stars. This may sound like a small amount of mass for an object as large as a galaxy, but only three new stars (out of billions in the *G*alaxy) are formed each year.

Do planets orbit other stars or is ours the only planetary system? In the past few decades, new technology has enabled us to answer that question by revealing nearly 3500 exoplanets in over 2600 planetary systems. Even before planets were detected, astronomers had predicted that planetary systems were likely to be byproducts of the star-formation process. In this chapter, we look at how interstellar matter is transformed into stars and planets.

Star Formation

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the sometimes-violent processes by which parts of a molecular cloud collapse to produce stars
- Recognize some of the structures seen in images of molecular clouds like the one in Orion
- Explain how the environment of a molecular cloud enables the formation of stars
- Describe how advancing waves of star formation cause a molecular cloud to evolve

As we begin our exploration of how stars are formed, let’s review some basics about stars discussed in earlier chapters:

- Stable (main-sequence) stars such as our Sun maintain equilibrium by producing energy through nuclear fusion in their cores. The ability to generate energy by fusion defines a star.
- Each second in the Sun, approximately 600 million tons of hydrogen undergo fusion into helium, with about 4 million tons turning into energy in the process. This rate of hydrogen use means that eventually the Sun (and all other stars) will run out of central fuel.
- Stars come with many different masses, ranging from 1/12 solar masses (M_{Sun}) to roughly 100–200 M_{Sun}. There are far more low-mass than high-mass stars.

- The most massive main-sequence stars (spectral type O) are also the most luminous and have the highest surface temperature. The lowest-mass stars on the main sequence (spectral type M or L) are the least luminous and the coolest.
- A galaxy of stars such as the Milky Way contains enormous amounts of gas and dust—enough to make billions of stars like the Sun.

If we want to find stars still in the process of formation, we must look in places that have plenty of the raw material from which stars are assembled. Since stars are made of gas, we focus our attention (and our telescopes) on the dense and cold clouds of gas and dust that dot the Milky Way (see [link] and [link]).

Pillars of Dust and Dense Globules in M16.

(a) (b)

(a) This Hubble Space Telescope image of the central regions of M16 (also known as the Eagle Nebula) shows huge columns of cool gas, (including molecular hydrogen, H2) and dust. These columns are of higher density than the surrounding regions and have resisted evaporation by the ultraviolet radiation from a cluster of hot stars just beyond the upper-right corner of this image. The tallest pillar is about 1 light-year long, and the M16 region is about 7000 light-years away from us. (b) This close-up view of one of the pillars shows some very dense globules, many of which harbor embryonic stars. Astronomers coined the term *evaporating gas globules* (EGGs) for these structures,

in part so that they could say we found EGGs inside the Eagle Nebula. It is possible that because these EGGs are exposed to the relentless action of the radiation from nearby hot stars, some may not yet have collected enough material to form a star. (credit a : modification of work by NASA, ESA, and the Hubble Heritage Team (STScI/AURA); credit b: modification of work by NASA, ESA, STScI, J. Hester and P. Scowen (Arizona State University))

Molecular Clouds: Stellar Nurseries

As we saw in Between the Stars: Gas and Dust in Space, the most massive reservoirs of interstellar matter—and some of the most massive objects in the Milky Way Galaxy—are the **giant molecular clouds**. These clouds have cold interiors with characteristic temperatures of only 10–20 K; most of their gas atoms are bound into molecules. These clouds turn out to be the birthplaces of most stars in our Galaxy.

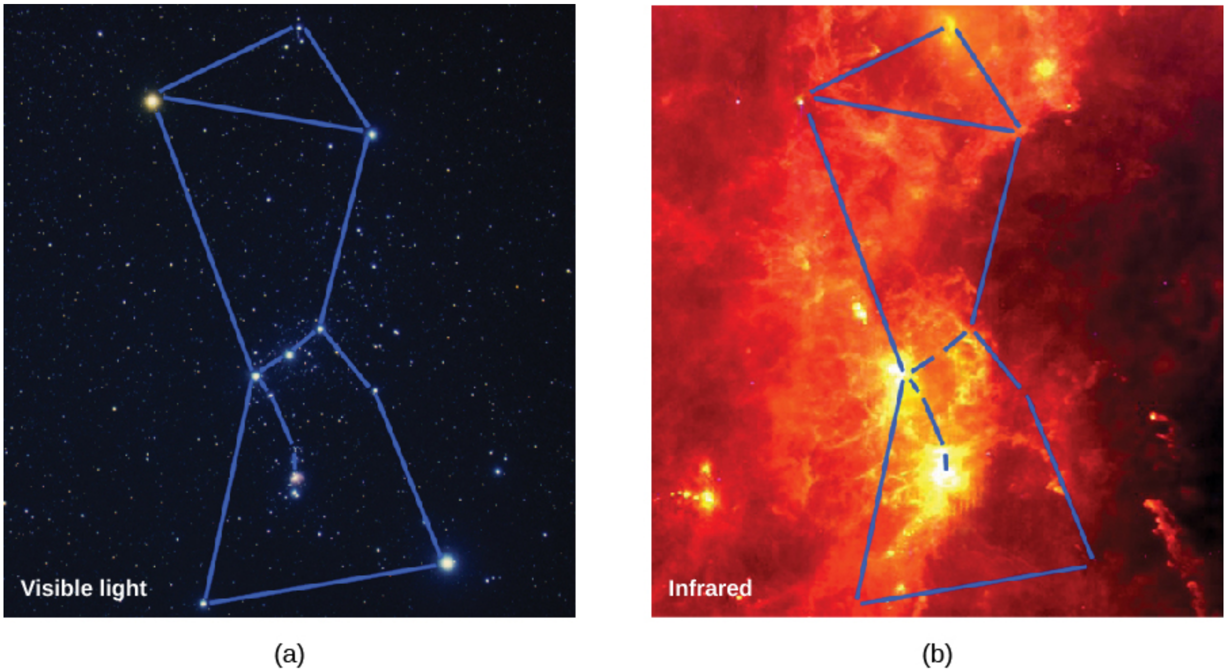
The masses of molecular clouds range from a thousand times the mass of the Sun to about 3 million solar masses. Molecular clouds have a complex filamentary structure, similar to cirrus clouds in Earth's atmosphere, but much less dense. The molecular cloud filaments can be up to 1000 light-years long. Within the clouds are cold, dense regions with typical masses of 50 to 500 times the mass of the Sun; we give these regions the highly technical name *clumps*. Within these clumps, there are even denser, smaller regions called cores. The cores are the embryos of stars. The conditions in these cores—low temperature and high density—are just what is required to make stars. Remember that the essence of the life story of any star is the ongoing competition between two forces: *gravity* and *pressure*. The force of gravity, pulling inward, tries to make a star collapse. Internal pressure produced by the motions of the gas atoms, pushing outward, tries to force the star to expand. When a star is first forming, low temperature (and hence, low pressure) and high density (hence, greater gravitational attraction) both work to give gravity the advantage. In order to form a star—that is, a dense, hot ball of matter capable of starting nuclear reactions deep within—we need a typical core of interstellar atoms and molecules to shrink in radius

and increase in density by a factor of nearly 10^{20}. It is the force of gravity that produces this drastic collapse.

The Orion Molecular Cloud

Let's discuss what happens in regions of star formation by considering a nearby site where stars are forming right now. One of the best-studied stellar nurseries is in the constellation of Orion, The Hunter, about 1500 light-years away ([link]). The pattern of the hunter is easy to recognize by the conspicuous "belt" of three stars that mark his waist. The Orion molecular cloud is much larger than the star pattern and is truly an impressive structure. In its long dimension, it stretches over a distance of about 100 light-years. The total quantity of molecular gas is about 200,000 times the mass of the Sun. Most of the cloud does not glow with visible light but betrays its presence by the radiation that the dusty gas gives off at infrared and radio wavelengths.

Orion in Visible and Infrared.

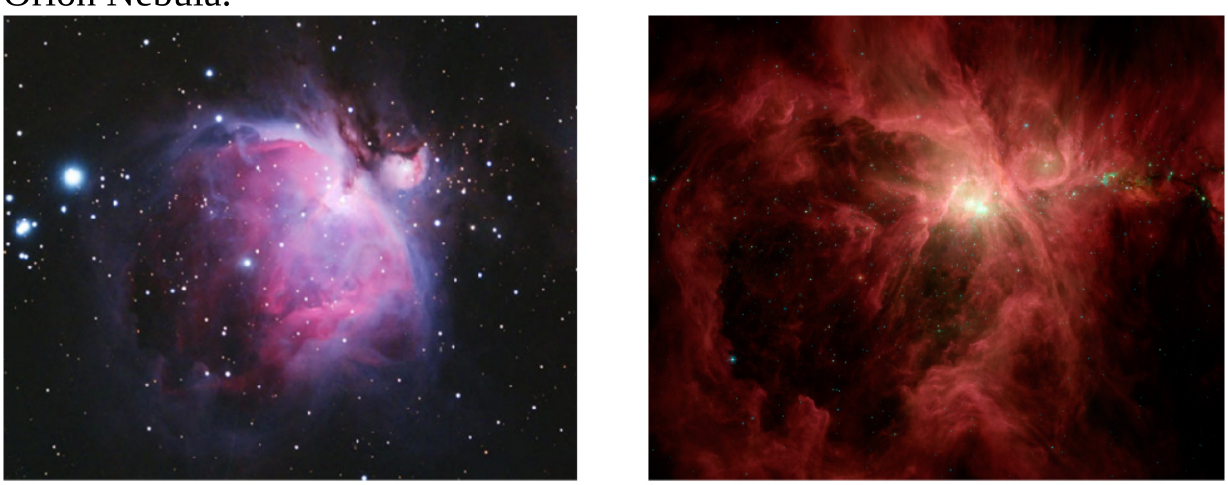


(a) The Orion star group was named after the legendary hunter in Greek mythology. Three stars close together in a link mark Orion's belt. The ancients imagined a sword hanging from the belt; the object at the end of the blue line in this sword is the Orion Nebula. (b) This

wide-angle, infrared view of the same area was taken with the Infrared Astronomical Satellite. Heated dust clouds dominate in this false-color image, and many of the stars that stood out on part (a) are now invisible. An exception is the cool, red-giant star Betelgeuse, which can be seen as a yellowish point at the left vertex of the blue triangle (at Orion’s left armpit). The large, yellow ring to the right of Betelgeuse is the remnant of an exploded star. The infrared image lets us see how large and full of cooler material the Orion molecular cloud really is. On the visible-light image at left, you see only two colorful regions of interstellar matter—the two, bright yellow splotches at the left end of and below Orion’s belt. The lower one is the Orion Nebula and the higher one is the region of the Horsehead Nebula. (credit: modification of work by NASA, visible light: Akira Fujii; infrared: Infrared Astronomical Satellite)

The stars in Orion’s belt are typically about 5 million years old, whereas the stars near the middle of the “sword” hanging from Orion’s belt are only 300,000 to 1 million years old. The region about halfway down the sword where star formation is still taking place is called the Orion Nebula. About 2200 young stars are found in this region, which is only slightly larger than a dozen light-years in diameter. The Orion Nebula also contains a tight cluster of stars called the Trapezium ([link]). The brightest Trapezium stars can be seen easily with a small telescope.

Orion Nebula.

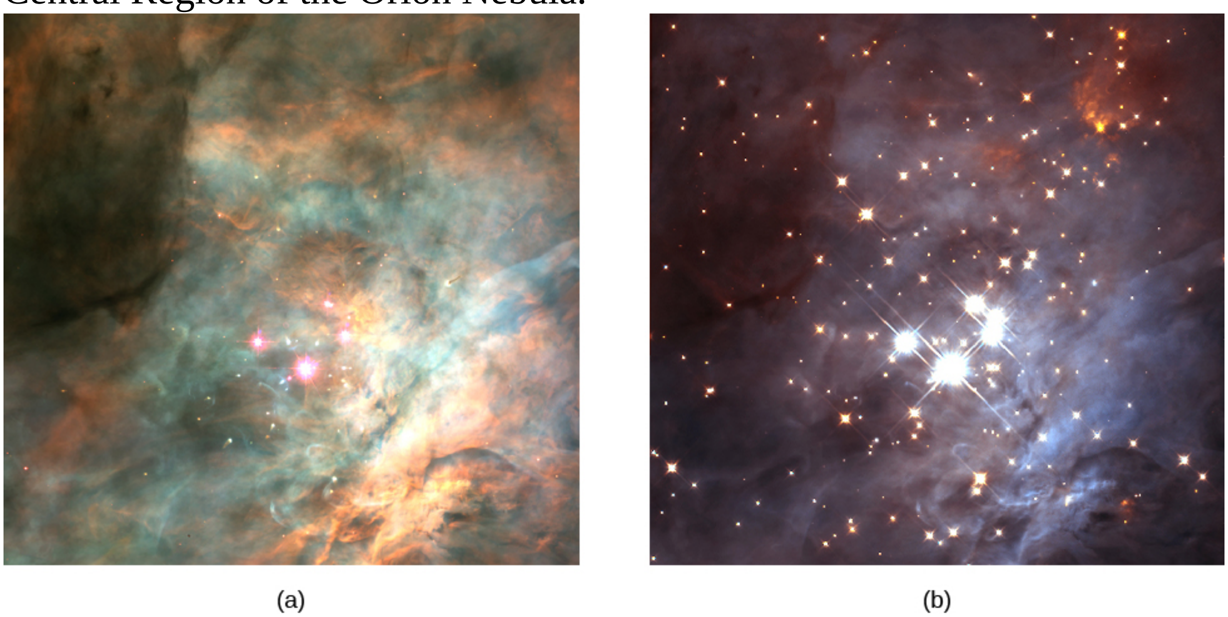


(a) (b)

(a) The Orion Nebula is shown in visible light. (b) With near-infrared radiation, we can see more detail within the dusty nebula since infrared can penetrate dust more easily than can visible light. (credit a: modification of work by Filip Lolić; credit b: modification of work by NASA/JPL-Caltech/T. Megeath (University of Toledo, Ohio))

Compare this with our own solar neighborhood, where the typical spacing between stars is about 3 light-years. Only a small number of stars in the Orion cluster can be seen with visible light, but infrared images—which penetrate the dust better—detect the more than 2000 stars that are part of the group ([link]).

Central Region of the Orion Nebula.



The Orion Nebula harbors some of the youngest stars in the solar neighborhood. At the heart of the nebula is the Trapezium cluster, which includes four very bright stars that provide much of the energy that causes the nebula to glow so brightly. In these images, we see a section of the nebula in (a) visible light and (b) infrared. The four bright stars in the center of the visible-light image are the Trapezium stars. Notice that most of the stars seen in the infrared are completely hidden by dust in the visible-light image. (credit a: modification of

work by NASA, C.R. O’Dell and S.K. Wong (Rice University); credit b: modification of work by NASA; K.L. Luhman (Harvard-Smithsonian Center for Astrophysics); and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory, University of Arizona))

Studies of Orion and other star-forming regions show that star formation is not a very efficient process. In the region of the Orion Nebula, about 1% of the material in the cloud has been turned into stars. That is why we still see a substantial amount of gas and dust near the Trapezium stars. The leftover material is eventually heated, either by the radiation and winds from the hot stars that form or by explosions of the most massive stars. (We will see in later chapters that the most massive stars go through their lives very quickly and end by exploding.)

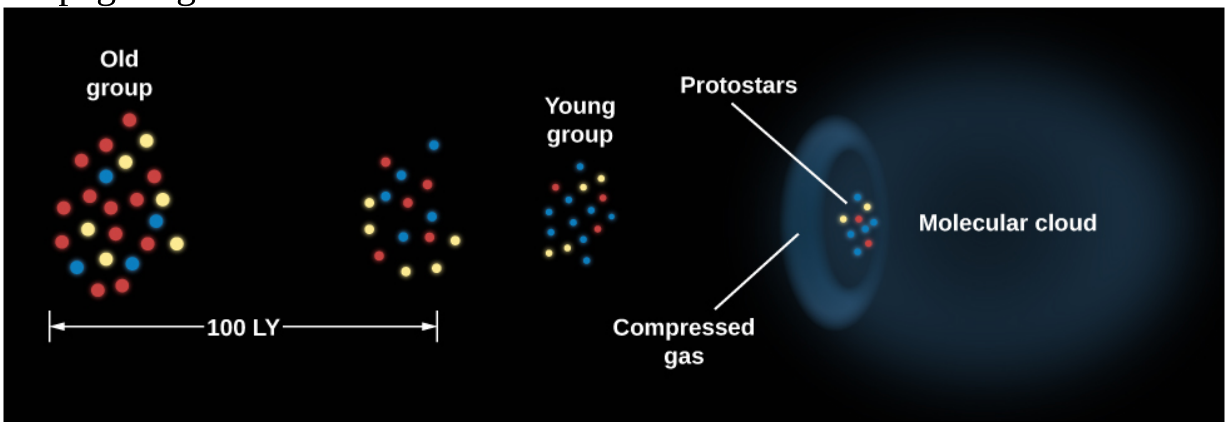
Note:
Take a journey through the Orion Nebula to view a nice narrated video tour of this region.

Whether gently or explosively, the material in the neighborhood of the new stars is blown away into interstellar space. Older groups or clusters of stars can now be easily observed in visible light because they are no longer shrouded in dust and gas ([link]).
Westerlund 2.

This young cluster of stars known as Westerlund 2 formed within the Carina star-forming region about 2 million years ago. Stellar winds and pressure produced by the radiation from the hot stars within the cluster are blowing and sculpting the surrounding gas and dust. The nebula still contains many globules of dust. Stars are continuing to form within the denser globules and pillars of the nebula. This Hubble Space Telescope image includes near-infrared exposures of the star cluster and visible-light observations of the surrounding nebula. Colors in the nebula are dominated by the red glow of hydrogen gas, and blue-green emissions from glowing oxygen. (credit: NASA, ESA, the Hubble Heritage Team (STScI/AURA), A. Nota (ESA/STScI), and the Westerlund 2 Science Team)

Although we do not know what initially caused stars to begin forming in Orion, there is good evidence that the first generation of stars triggered the formation of additional stars, which in turn led to the formation of still more stars ([link]).

Propagating Star Formation.



Star formation can move progressively through a molecular cloud. The oldest group of stars lies to the left of the diagram and has expanded because of the motions of individual stars. Eventually, the stars in the group will disperse and no longer be recognizable as a cluster. The youngest group of stars lies to the right, next to the molecular cloud. This group of stars is only 1 to 2 million years old. The pressure of the hot, ionized gas surrounding these stars compresses the material in the nearby edge of the molecular cloud and initiates the gravitational collapse that will lead to the formation of more stars.

The basic idea of triggered star formation is this: when a massive star is formed, it emits a large amount of ultraviolet radiation and ejects high-speed gas in the form of a stellar wind. This injection of energy heats the gas around the stars and causes it to expand. When massive stars exhaust their supply of fuel, they explode, and the energy of the explosion also heats the gas. The hot gases pile into the surrounding cold molecular cloud, compressing the material in it and increasing its density. If this increase in density is large enough, gravity will overcome pressure, and stars will begin to form in the compressed gas. Such a chain reaction—where the brightest and hottest stars of one area become the cause of star formation “next

door”—seems to have occurred not only in Orion but also in many other molecular clouds.

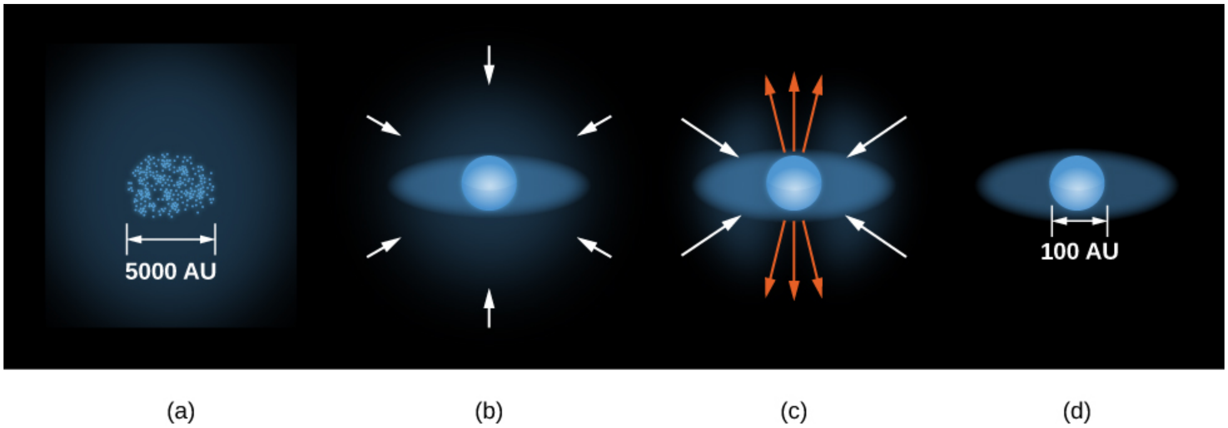
There are many molecular clouds that form only (or mainly) low-mass stars. Because low-mass stars do not have strong winds and do not die by exploding, triggered star formation cannot occur in these clouds. There are also stars that form in relative isolation in small cores. Therefore, not all star formation is originally triggered by the death of massive stars. However, there are likely to be other possible triggers, such as spiral density waves and other processes we do not yet understand.

The Birth of a Star

Although regions such as Orion give us clues about how star formation begins, the subsequent stages are still shrouded in mystery (and a lot of dust). There is an enormous difference between the density of a molecular cloud core and the density of the youngest stars that can be detected. Direct observations of this collapse to higher density are nearly impossible for two reasons. First, the dust-shrouded interiors of molecular clouds where stellar births take place cannot be observed with visible light. Second, the timescale for the initial collapse—thousands of years—is very short, astronomically speaking. Since each star spends such a tiny fraction of its life in this stage, relatively few stars are going through the collapse process at any given time. Nevertheless, through a combination of theoretical calculations and the limited observations available, astronomers have pieced together a picture of what the earliest stages of stellar evolution are likely to be.

The first step in the process of creating stars is the formation of dense cores within a clump of gas and dust ([link](a)). It is generally thought that all the material for the star comes from the core, the larger structure surrounding the forming star. Eventually, the gravitational force of the infalling gas becomes strong enough to overwhelm the pressure exerted by the cold material that forms the dense cores. The material then undergoes a rapid collapse, and the density of the core increases greatly as a result. During the time a dense core is contracting to become a true star, but before the fusion of protons to produce helium begins, we call the object a **protostar**.

Formation of a Star.



(a) Dense cores form within a molecular cloud. (b) A protostar with a surrounding disk of material forms at the center of a dense core, accumulating additional material from the molecular cloud through gravitational attraction. (c) A stellar wind breaks out but is confined by the disk to flow out along the two poles of the star. (d) Eventually, this wind sweeps away the cloud material and halts the accumulation of additional material, and a newly formed star, surrounded by a disk, becomes observable. These sketches are not drawn to the same scale. The diameter of a typical envelope that is supplying gas to the newly forming star is about 5000 AU. The typical diameter of the disk is about 100 AU or slightly larger than the diameter of the orbit of Pluto.

The natural turbulence inside a clump tends to give any portion of it some initial spinning motion (even if it is very slow). As a result, each collapsing core is expected to spin. According to the law of conservation of angular momentum (discussed in the chapter on [Orbits and Gravity](#)), a rotating body spins more rapidly as it decreases in size. In other words, if the object can turn its material around a smaller circle, it can move that material more quickly—like a figure skater spinning more rapidly as she brings her arms in tight to her body. This is exactly what happens when a core contracts to form a protostar: as it shrinks, its rate of spin increases.

But all directions on a spinning sphere are not created equal. As the protostar rotates, it is much easier for material to fall right onto the poles

(which spin most slowly) than onto the equator (where material moves around most rapidly). Therefore, gas and dust falling in toward the protostar's equator are "held back" by the rotation and form a whirling extended disk around the equator (part b in [link]). You may have observed this same "equator effect" on the amusement park ride in which you stand with your back to a cylinder that is spun faster and faster. As you spin really fast, you are pushed against the wall so strongly that you cannot possibly fall toward the center of the cylinder. Gas can, however, fall onto the protostar easily from directions away from the star's equator.

The protostar and disk at this stage are embedded in an envelope of dust and gas from which material is still falling onto the protostar. This dusty envelope blocks visible light, but infrared radiation can get through. As a result, in this phase of its evolution, the protostar itself is emitting infrared radiation and so is observable only in the infrared region of the spectrum. Once almost all of the available material has been accreted and the central protostar has reached nearly its final mass, it is given a special name: it is called a *T Tauri star*, named after one of the best studied and brightest members of this class of stars, which was discovered in the constellation of Taurus. (Astronomers have a tendency to name types of stars after the first example they discover or come to understand. It's not an elegant system, but it works.) Only stars with masses less than or similar to the mass of the Sun become T Tauri stars. Massive stars do not go through this stage, although they do appear to follow the formation scenario illustrated in [link].

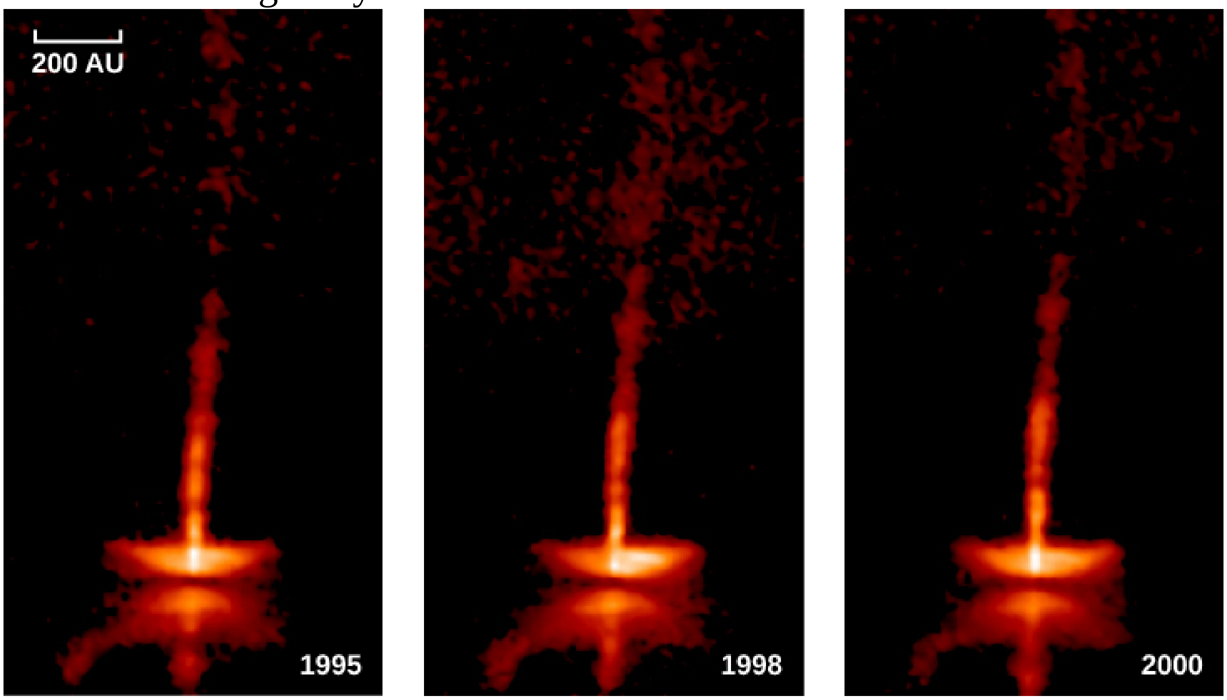
Winds and Jets

Recent observations suggest that T Tauri stars may actually be stars in a middle stage between protostars and hydrogen-fusing stars such as the Sun. High-resolution infrared images have revealed jets of material as well as *stellar winds* coming from some T Tauri stars, proof of interaction with their environment. A **stellar wind** consists mainly of protons (hydrogen nuclei) and electrons streaming away from the star at speeds of a few hundred kilometers per second (several hundred thousand miles per hour). When the wind first starts up, the disk of material around the star's equator

blocks the wind in its direction. Where the wind particles *can* escape most effectively is in the direction of the star’s poles.

Astronomers have actually seen evidence of these beams of particles shooting out in opposite directions from the polar regions of newly formed stars. In many cases, these beams point back to the location of a protostar that is still so completely shrouded in dust that we cannot yet see it ([link]).

Gas Jets Flowing away from a Protostar.

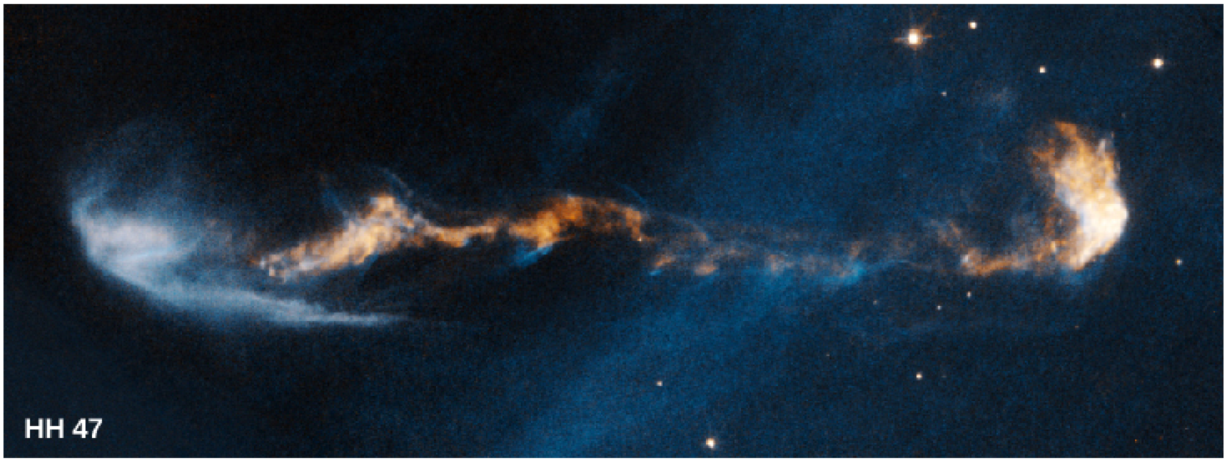


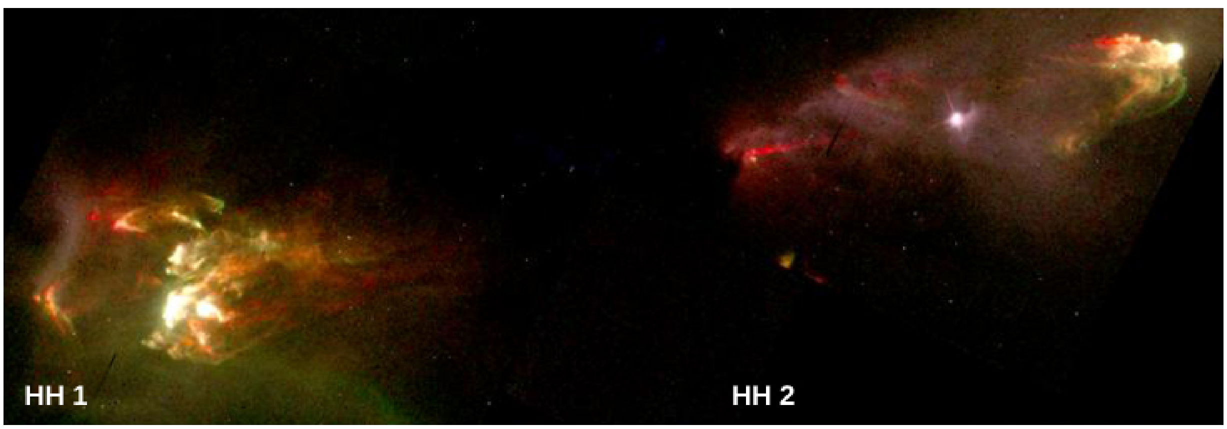
Here we see the neighborhood of a protostar, known to us as HH 34 because it is a Herbig-Haro object. The star is about 450 light-years away and only about 1 million years old. Light from the star itself is blocked by a disk, which is larger than 60 billion kilometers in diameter and is seen almost edge-on. Jets are seen emerging perpendicular to the disk. The material in these jets is flowing outward at speeds up to 580,000 kilometers per hour. The series of three images shows changes during a period of 5 years. Every few months, a compact clump of gas is ejected, and its motion outward can be followed. The changes in the brightness of the disk may be due to motions of clouds within the disk that alternately block some of the light and then let it through. This image corresponds to the stage in the

life of a protostar shown in part (c) of [link]. (credit: modification of work by Hubble Space Telescope, NASA, ESA)

On occasion, the jets of high-speed particles streaming away from the protostar collide with a somewhat-denser lump of gas nearby, excite its atoms, and cause them to emit light. These glowing regions, each of which is known as a **Herbig-Haro (HH) object** after the two astronomers who first identified them, allow us to trace the progress of the jet to a distance of a light-year or more from the star that produced it. [link] shows two spectacular images of HH objects.

Outflows from Protostars.



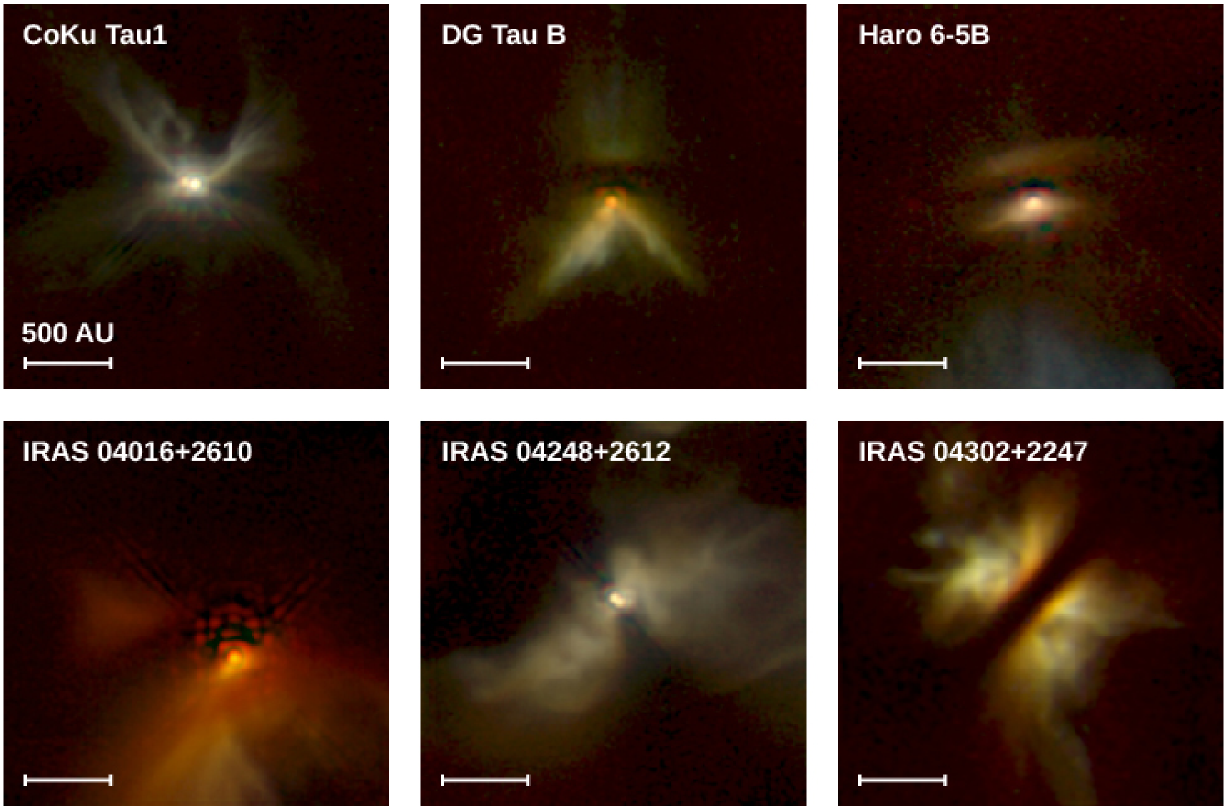


These images were taken with the Hubble Space Telescope and show jets flowing outward from newly formed stars. In the HH47 image, a protostar 1500 light-years away (invisible inside a dust disk at the left

edge of the image) produces a very complicated jet. The star may actually be wobbling, perhaps because it has a companion. Light from the star illuminates the white region at the left because light can emerge perpendicular to the disk (just as the jet does). At right, the jet is plowing into existing clumps of interstellar gas, producing a shock wave that resembles an arrowhead. The HH1/2 image shows a double-beam jet emanating from a protostar (hidden in a dust disk in the center) in the constellation of Orion. Tip to tip, these jets are more than 1 light-year long. The bright regions (first identified by Herbig and Haro) are places where the jet is a slamming into a clump of interstellar gas and causing it to glow. (credit "HH 47": modification of work by NASA, ESA, and P. Hartigan (Rice University); credit "HH 1 and HH 2: modification of work by J. Hester, WFPC2 Team, NASA)

The wind from a forming star will ultimately sweep away the material that remains in the obscuring envelope of dust and gas, leaving behind the naked disk and protostar, which can then be seen with visible light. We should note that at this point, the protostar itself is still contracting slowly and has not yet reached the main-sequence stage on the H–R diagram (a concept introduced in the chapter The Stars: A Celestial Census). The disk can be detected directly when observed at infrared wavelengths or when it is seen silhouetted against a bright background ([link]).

Disks around Protostars.



These Hubble Space Telescope infrared images show disks around young stars in the constellation of Taurus, in a region about 450 light-years away. In some cases, we can see the central star (or stars—some are binaries). In other cases, the dark, horizontal bands indicate regions where the dust disk is so thick that even infrared radiation from the star embedded within it cannot make its way through. The brightly glowing regions are starlight reflected from the upper and lower surfaces of the disk, which are less dense than the central, dark regions. (Credit: modification of work by D. Padgett (IPAC/Caltech), W. Brandner (IPAC), K. Stapelfeldt (JPL) and NASA)

This description of a protostar surrounded by a rotating disk of gas and dust sounds very much like what happened in our solar system when the Sun and planets formed. Indeed, one of the most important discoveries from the study of star formation in the last decade of the twentieth century was that disks are an inevitable byproduct of the process of creating stars. The next questions that astronomers set out to answer was: will the disks around

protostars also form planets? And if so, how often? We will return to these questions later in this chapter.

To keep things simple, we have described the formation of single stars. Many stars, however, are members of binary or triple systems, where several stars are born together. In this case, the stars form in nearly the same way. Widely separated binaries may each have their own disk; close binaries may share a single disk.

Key Concepts and Summary

Most stars form in giant molecular clouds with masses as large as 3×10^6 solar masses. The most well-studied molecular cloud is Orion, where star formation is currently taking place. Molecular clouds typically contain regions of higher density called clumps, which in turn contain several even-denser cores of gas and dust, each of which may become a star. A star can form inside a core if its density is high enough that gravity can overwhelm the internal pressure and cause the gas and dust to collapse. The accumulation of material halts when a protostar develops a strong stellar wind, leading to jets of material being observed coming from the star. These jets of material can collide with the material around the star and produce regions that emit light that are known as Herbig-Haro objects.

Glossary

giant molecular clouds
: large, cold interstellar clouds with diameters of dozens of light-years and typical masses of 10^5 solar masses; found in the spiral arms of galaxies, these clouds are where stars form

Herbig-Haro (HH) object
: luminous knots of gas in an area of star formation that are set to glow by jets of material from a protostar

protostar
: a very young star still in the process of formation, before nuclear fusion begins

stellar wind

the outflow of gas, sometimes at speeds as high as hundreds of kilometers per second, from a star

Evidence That Planets Form around Other Stars

> **LEARNING OBJECTIVES**
>
> By the end of this section, you will be able to:
>
> - Trace the evolution of dust surrounding a protostar, leading to the development of rocky planets and gas giants
> - Estimate the timescale for growth of planets using observations of the disks surrounding young stars
> - Evaluate evidence for planets around forming stars based on the structures seen in images of the circumstellar dust disks

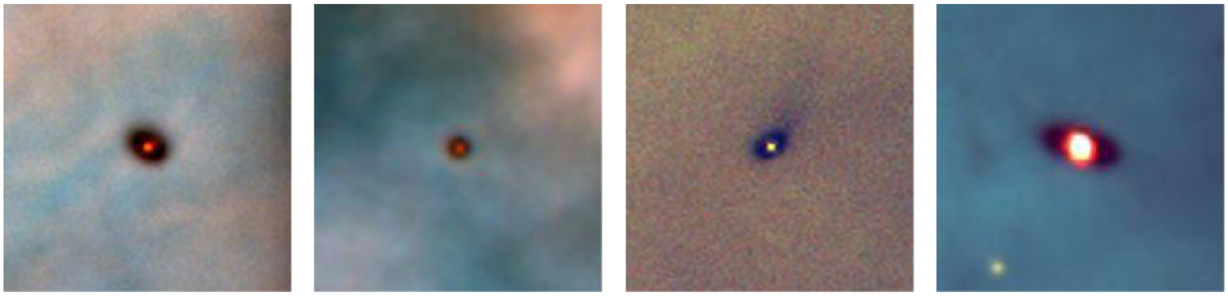
Having developed on a planet and finding it essential to our existence, we have a special interest in how planets fit into the story of star formation. Yet planets outside the solar system are extremely difficult to detect. Recall that we see planets in our own system only because they reflect sunlight and are close by. When we look to the other stars, we find that the amount of light a planet reflects is a depressingly tiny fraction of the light its star gives off. Furthermore, from a distance, planets are lost in the glare of their much-brighter parent stars.

Disks around Protostars: Planetary Systems in Formation

It is a lot easier to detect the spread-out raw material from which planets might be assembled than to detect planets after they are fully formed. From our study of the solar system, we understand that planets form by the gathering together of gas and dust particles in orbit around a newly created star. Each dust particle is heated by the young protostar and radiates in the

infrared region of the spectrum. Before any planets form, we can detect such radiation from all of the spread-out individual dust particles that are destined to become parts of planets. We can also detect the silhouette of the disk if it blocks bright light coming from a source behind it ([link]).

Disks around Protostars.



These Hubble Space Telescope images show four disks around young stars in the Orion Nebula. The dark, dusty disks are seen silhouetted against the bright backdrop of the glowing gas in the nebula. The size of each image is about 30 times the diameter of our planetary system; this means the disks we see here range in size from two to eight times the orbit of Pluto. The red glow at the center of each disk is a young star, no more than a million years old. These images correspond to the stage in the life of a protostar shown in part (d) of [link]. (credit: modification of work by Mark McCaughrean (Max-Planck-Institute for Astronomy), C. Robert O'Dell (Rice University), and NASA)

Once the dust particles gather together and form a few planets (and maybe some moons), the overwhelming majority of the dust is hidden in the interiors of the planets where we cannot see it. All we can now detect is the radiation from the outside surfaces, which cover a drastically smaller area than the huge, dusty disk from which they formed. The amount of infrared radiation is therefore greatest before the dust particles combine into planets. For this reason, our search for planets begins with a search for infrared radiation from the material required to make them.

A disk of gas and dust appears to be an essential part of star formation. Observations show that nearly all very young protostars have disks and that the disks range in size from 10 to 1000 AU. (For comparison, the average

diameter of the orbit of Pluto, which can be considered the rough size of our own planetary system, is 80 AU, whereas the outer diameter of the Kuiper belt of smaller icy bodies is about 100 AU.) The mass contained in these disks is typically 1–10% of the mass of our own Sun, which is more than the mass of all the planets in our solar system put together. Such observations already demonstrate that a large fraction of stars begin their lives with enough material in the right place to form a planetary system.

The Timing of Planet Formation and Growth

We can use observations of how the disks change with time to estimate how long it takes for planets to form. If we measure the temperature and luminosity of a protostar, then, as we saw, we can place it in an H–R diagram like the one shown in [link]. By comparing the real star with our models of how protostars should evolve with time, we can estimate its age. We can then look at how the disks we observe change with the ages of the stars that they surround.

What such observations show is that if a protostar is less than about 1 to 3 million years old, its disk extends all the way from very close to the surface of the star out to tens or hundreds of AU away. In older protostars, we find disks with outer parts that still contain large amounts of dust, but the inner regions have lost most of their dust. In these objects, the disk looks like a donut, with the protostar centered in its hole. The inner, dense parts of most disks have disappeared by the time the stars are 10 million years old ([link]).

Protoplanetary Disks around Two Stars.

The left view of each star shows infrared observations by the Hubble

Space Telescope of their protoplanetary disks. The central star is much brighter than the surrounding disk, so the instrument includes a coronograph, which has a small shield that blocks the light of the central star but allows the surrounding disk to be imaged. The right image of each star shows models of the disks based on the observations. The star HD 141943 has an age of about 17 million years, while HD 191089 is about 12 million years old. (credit: modification of work by NASA, ESA, R. Soummer and M. Perrin (STScI), L. Pueyo (STScI/Johns Hopkins University), C. Chen and D. Golimowski (STScI), J.B. Hagan (STScI/Purdue University), T. Mittal (University of California, Berkeley/Johns Hopkins University), E. Choquet, M. Moerchen, and M. N'Diaye (STScI), A. Rajan (Arizona State University), S. Wolff (STScI/Purdue University), J. Debes and D. Hines (STScI), and G. Schneider (Steward Observatory/University of Arizona))

Calculations show that the formation of one or more planets could produce such a donut-like distribution of dust. Suppose a planet forms a few AU away from the protostar, presumably due to the gathering together of matter from the disk. As the planet grows in mass, the process clears out a dust-free region in its immediate neighborhood. Calculations also show that any small dust particles and gas that were initially located in the region between the protostar and the planet, and that are not swept up by the planet, will then fall onto the star very quickly in about 50,000 years.

Matter lying outside the planet's orbit, in contrast, is prevented from moving into the hole by the gravitational forces exerted by the planet. (We saw something similar in Saturn's rings, where the action of the shepherd moons keeps the material near the edge of the rings from spreading out.) If the formation of a planet is indeed what produces and sustains holes in the disks that surround very young stars, then planets must form in 3 to 30 million years. This is a short period compared with the lifetimes of most stars and shows that the formation of planets may be a quick byproduct of the birth of stars.

Calculations show that *accretion* can drive the rapid growth of planets—small, dust-grain-size particles orbiting in the disk collide and stick together, with the larger collections growing more rapidly as they attract and capture smaller ones. Once these clumps grow to about 10 centimeters in size or so, they enter a perilous stage in their development. At that size, unless they can grow to larger than about 100 meters in diameter, they are subject to drag forces produced by friction with the gas in the disk—and their orbits can rapidly decay, plunging them into the host star. Therefore, these bodies must rapidly grow to nearly 1 kilometer in size in diameter to avoid a fiery fate. At this stage, they are considered planetesimals (the small chunks of solid matter—ice and dust particles—that you learned about in Other Worlds: An Introduction to the Solar System). Once they survive to those sizes, the largest survivors will continue to grow by accreting smaller planetesimals; ultimately, this process results in a few large planets.

If the growing planets reach a mass bigger than about 10 times the mass of Earth, their gravity is strong enough to capture and hold on to hydrogen gas that remains in the disk. At that point, they will grow in mass and radius rapidly, reaching giant planet dimensions. However, to do so requires that the rapidly evolving central star hasn't yet driven away the gas in the disk with its increasingly vigorous wind (see the earlier section on Star Formation). From observations, we see that the disk can be blown away within 10 million years, so growth of a giant planet must also be a very fast process, astronomically speaking.

Debris Disks and Shepherd Planets

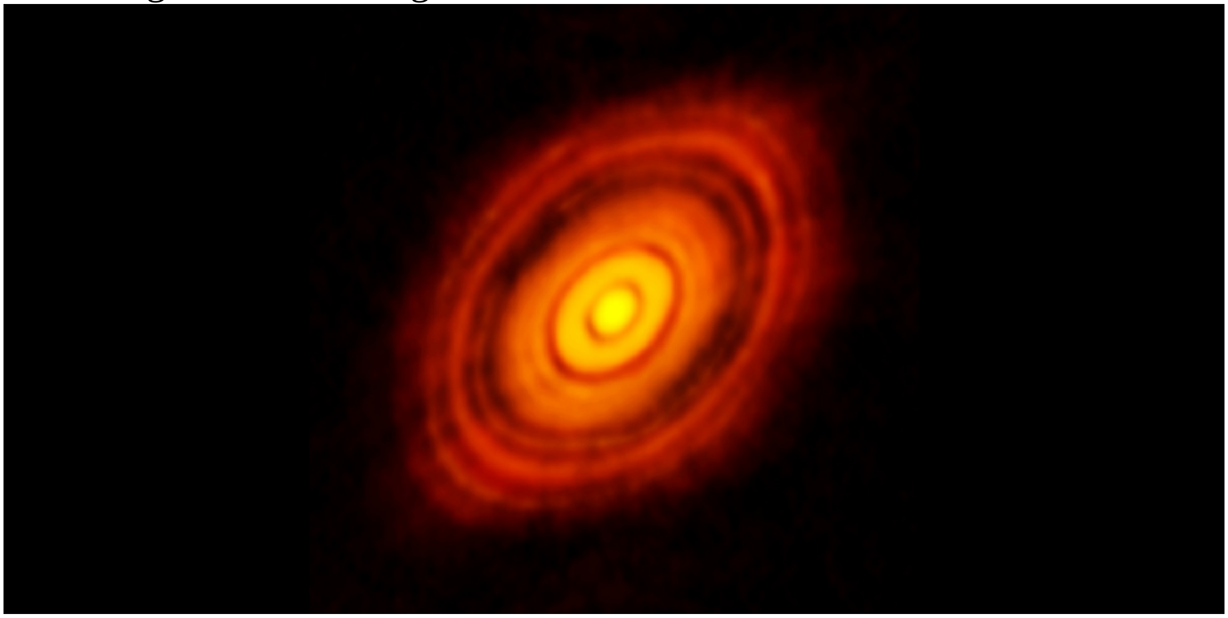
The dust around newly formed stars is gradually either incorporated into the growing planets in the newly forming planetary system or ejected through gravitational interactions with the planets into space. The dust will disappear after about 30 million years unless the disk is continually supplied with new material. Local comets and asteroids are the most likely sources of new dust. As the planet-size bodies grow, they stir up the orbits of smaller objects in the area. These small bodies collide at high speeds, shatter, and produce tiny particles of silicate dust and ices that can keep the disk supplied with the debris from these collisions.

Over several hundred million years, the comets and asteroids will gradually be reduced in number, the frequency of collisions will go down, and the supply of fresh dust will diminish. Remember that the heavy bombardment in the early solar system ended when the Sun was only about 500 million years old. Observations show that the dusty “debris disks” around stars also become largely undetectable by the time the stars reach an age of 400 to 500 million years. It is likely, however, that some small amount of cometary material will remain in orbit, much like our Kuiper belt, a flattened disk of comets outside the orbit of Neptune.

In a young planetary system, even if we cannot see the planets directly, the planets can concentrate the dust particles into clumps and arcs that are much larger than the planets themselves and more easily imaged. This is similar to how the tiny moons of Saturn shepherd the particles in the rings and produce large arcs and structures in Saturn’s rings.

Debris disks—many with just such clumps and arcs—have now been found around many stars, such as HL Tau, located about 450 light-years from Earth in the constellation Taurus ([link]). In some stars, the brightness of the rings varies with position; around other stars, there are bright arcs and gaps in the rings. The brightness indicates the relative concentration of dust, since what we are seeing is infrared (heat radiation) from the dust particles in the rings. More dust means more radiation.

Dust Ring around a Young Star.



This image was made by ALMA (the Atacama Large Millimeter/Submillimeter Array) at a wavelength of 1.3 millimeters and shows the young star HL Tau and its protoplanetary disk. It reveals multiple rings and gaps that indicate the presence of emerging planets, which are sweeping their orbits clear of dust and gas. (credit: modification of work by ALMA (ESO/NAOJ/NRAO))

Note:
Watch a short video clip of the director of NRAO (National Radio Astronomy Observatory) describing the high-resolution observations of the young star HL Tau. While you're there, watch an artist's animation of a protoplanetary disk to see newly formed planets traveling around a host (parent) star.

Key Concepts and Summary

Observational evidence shows that most protostars are surrounded by disks with large-enough diameters and enough mass (as much as 10% that of the Sun) to form planets. After a few million years, the inner part of the disk is cleared of dust, and the disk is then shaped like a donut with the protostar centered in the hole—something that can be explained by the formation of planets in that inner zone. Around a few older stars, we see disks formed from the debris produced when small bodies (comets and asteroids) collide with each other. The distribution of material in the rings of debris disks is probably determined by shepherd planets, just as Saturn's shepherd moons affect the orbits of the material in its rings. Protoplanets that grow to be 10 times the mass of Earth or bigger while there is still considerable gas in their disk can then capture more of that gas and become giant planets like Jupiter in the solar system.

Planets beyond the Solar System: Search and Discovery

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the orbital motion of planets in our solar system using Kepler’s laws
- Compare the indirect and direct observational techniques for exoplanet detection

For centuries, astronomers have dreamed of finding planets around other stars, including other planets like Earth. Direct observations of such distant planets are very difficult, however. You might compare a planet orbiting a star to a mosquito flying around one of those giant spotlights at a shopping center opening. From close up, you might spot the mosquito. But imagine viewing the scene from some distance away—say, from an airplane. You could see the spotlight just fine, but what are your chances of catching the mosquito in that light? Instead of making direct images, astronomers have relied on indirect observations and have now succeeded in detecting a multitude of planets around other stars.

In 1995, after decades of effort, we found the first such **exoplanet** (a planet outside our solar system) orbiting a main-sequence star, and today we know that most stars form with planets. This is an example of how persistence and new methods of observation advance the knowledge of humanity. By studying exoplanets, astronomers hope to better understand our solar system in context of the rest of the universe. For instance, how does the arrangement of our solar system compare to planetary systems in the rest of

the universe? What do exoplanets tell us about the process of planet formation? And how does knowing the frequency of exoplanets influence our estimates of whether there is life elsewhere?

Searching for Orbital Motion

Most exoplanet detections are made using techniques where we observe the *effect* that the planet exerts on the host star. For example, the gravitational tug of an unseen planet will cause a small wobble in the host star. Or, if its orbit is properly aligned, a planet will periodically cross in front of the star, causing the brightness of the star to dim.

To understand how a planet can move its host star, consider a single Jupiter-like planet. Both the planet and the star actually revolve about their *common center of mass*. Remember that gravity is a mutual attraction. The star and the planet each exert a force on the other, and we can find a stable point, the center of mass, between them about which both objects move. The smaller the mass of a body in such a system, the larger its orbit. A massive star barely swings around the center of mass, while a low-mass planet makes a much larger “tour.”

Suppose the planet is like Jupiter and has a mass about one-thousandth that of its star; in this case, the size of the star’s orbit is one-thousandth the size of the planet’s. To get a sense of how difficult observing such motion might be, let’s see how hard Jupiter would be to detect in this way from the distance of a nearby star. Consider an alien astronomer trying to observe our own system from Alpha Centauri, the closest star system to our own (about 4.3 light-years away). There are two ways this astronomer could try to detect the orbital motion of the Sun. One way would be to look for changes in the Sun’s position on the sky. The second would be to use the Doppler effect to look for changes in its velocity. Let’s discuss each of these in turn.

The diameter of Jupiter’s apparent orbit viewed from Alpha Centauri is 10 seconds of arc, and that of the Sun’s orbit is 0.010 seconds of arc. (Remember, 1 second of arc is 1/3600 degree.) If they could measure the apparent position of the Sun (which is bright and easy to detect) to

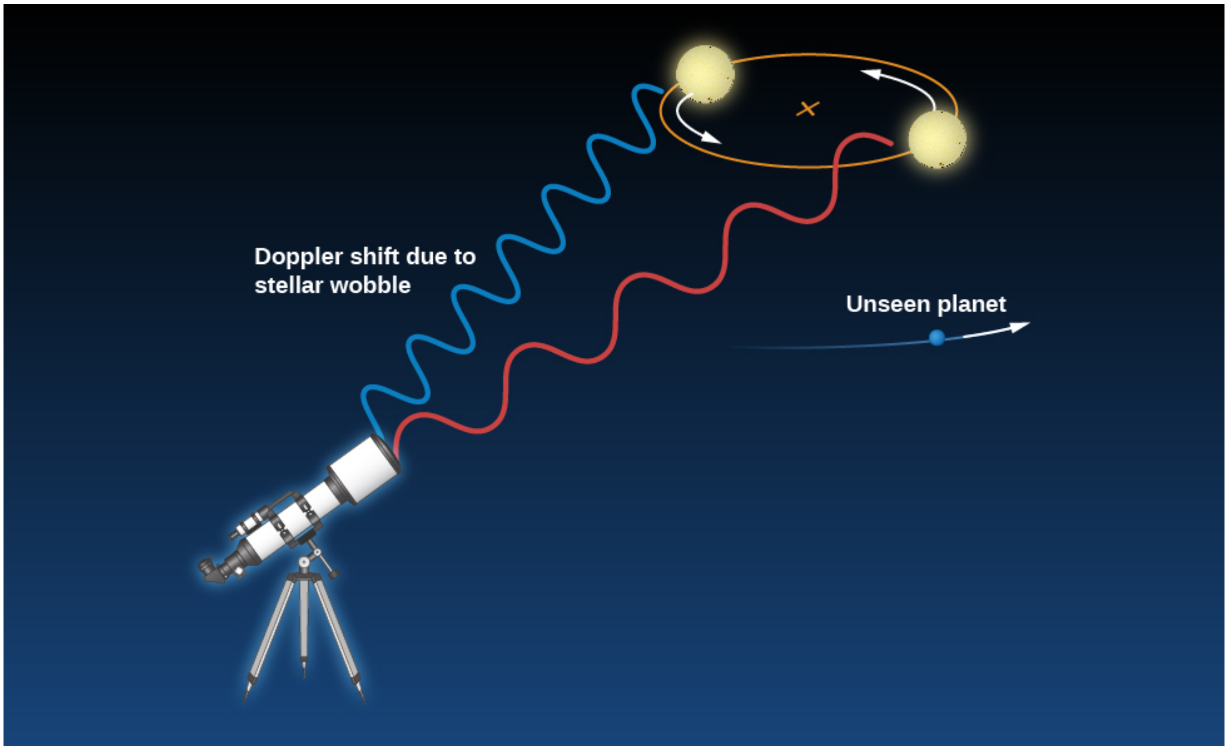
sufficient precision, they would describe an orbit of diameter 0.010 seconds of arc with a period equal to that of Jupiter, which is 12 years.

In other words, if they watched the Sun for 12 years, they would see it wiggle back and forth in the sky by this minuscule fraction of a degree. From the observed motion and the period of the “wiggle,” they could deduce the mass of Jupiter and its distance using Kepler’s laws. (To refresh your memory about these laws, see the chapter on Orbits and Gravity.)

Measuring positions in the sky this accurately is extremely difficult, and so far, astronomers have not made any confirmed detections of planets using this technique. However, we have been successful in using spectrometers to measure the changing velocity of stars with planets around them.

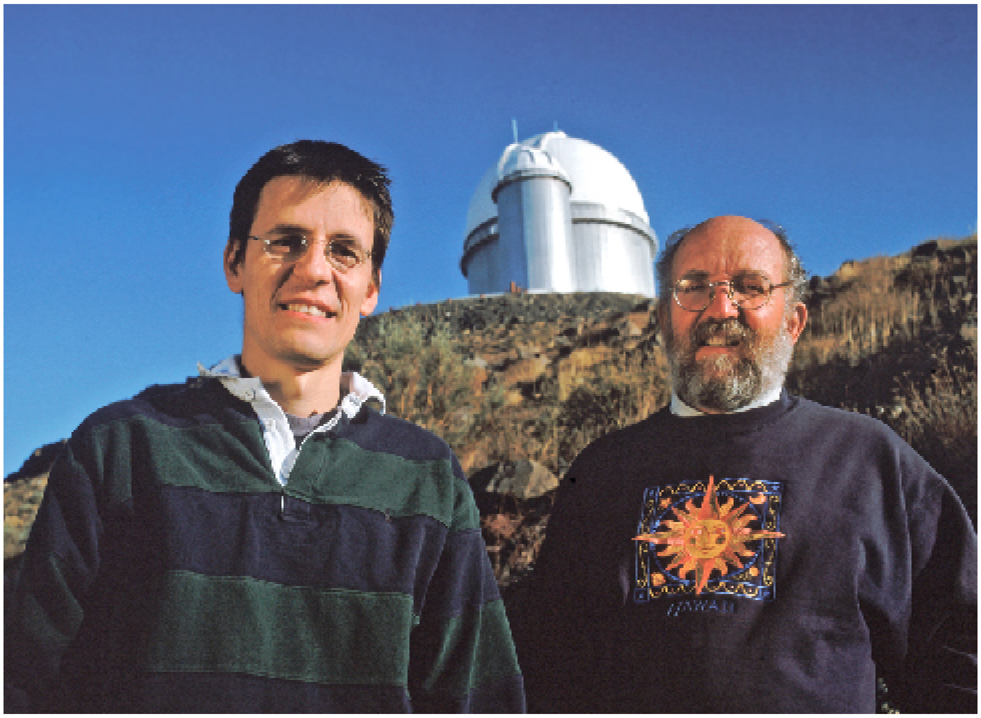
As the star and planet orbit each other, part of their motion will be in our line of sight (toward us or away from us). Such motion can be measured using the *Doppler effect* and the star’s spectrum. As the star moves back and forth in orbit around the system’s center of mass in response to the gravitational tug of an orbiting planet, the lines in its spectrum will shift back and forth.

Let’s again consider the example of the Sun. Its *radial velocity* (motion toward or away from us) changes by about 13 meters per second with a period of 12 years because of the gravitational pull of Jupiter. This corresponds to about 30 miles per hour, roughly the speed at which many of us drive around town. Detecting motion at this level in a star’s spectrum presents an enormous technical challenge, but several groups of astronomers around the world, using specialized spectrographs designed for this purpose, have succeeded. Note that the change in speed does not depend on the distance of the star from the observer. Using the Doppler effect to detect planets will work at any distance, as long as the star is bright enough to provide a good spectrum and a large telescope is available to make the observations ([link]).
Doppler Method of Detecting Planets.



The motion of a star around a common center of mass with an orbiting planet can be detected by measuring the changing speed of the star. When the star is moving away from us, the lines in its spectrum show a tiny redshift; when it is moving toward us, they show a tiny blueshift. The change in color (wavelength) has been exaggerated here for illustrative purposes. In reality, the Doppler shifts we measure are extremely small and require sophisticated equipment to be detected.

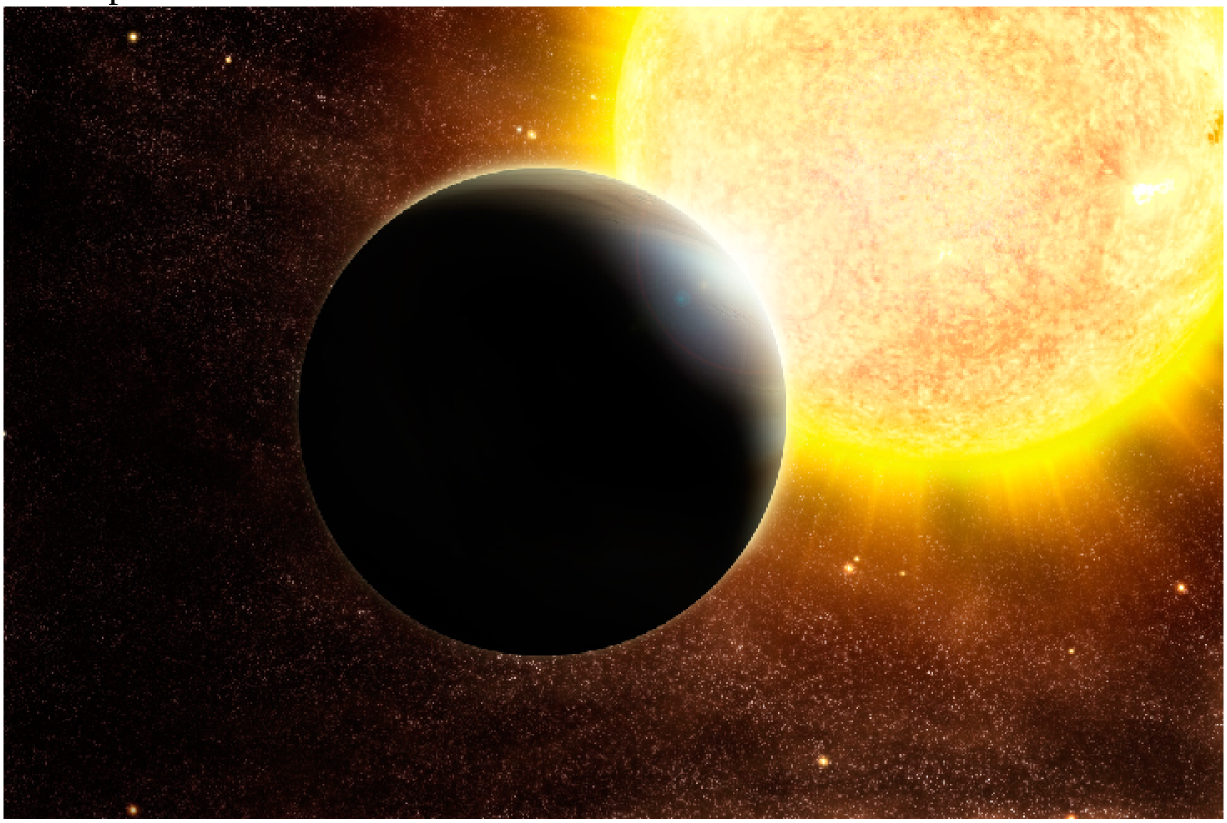
The first successful use of the Doppler effect to find a planet around another star was in 1995. Michel Mayor and Didier Queloz of the Geneva Observatory ([link]) used this technique to find a planet orbiting a star resembling our Sun called 51 Pegasi, about 40 light-years away. (The star can be found in the sky near the great square of Pegasus, the flying horse of Greek mythology, one of the easiest-to-find star patterns.) To everyone's surprise, the planet takes a mere 4.2 days to orbit around the star. (Remember that Mercury, the innermost planet in our solar system, takes 88 days to go once around the Sun, so 4.2 days seems fantastically short.)
Planet Discoverers.



In 1995, Didier Queloz and Michel Mayor of the Geneva Observatory were the first to discover a planet around a regular star (51 Pegasi), for which they received the 2019 Nobel Prize in physics. They are seen here at an observatory in Chile where they are continuing their planet hunting. (credit: Weinstein/Ciel et Espace Photos)

Mayor and Queloz’s findings mean the planet must be very close to 51 Pegasi, circling it about 7 million kilometers away ([link]). At that distance, the energy of the star should heat the planet’s surface to a temperature of a few thousand degrees Celsius (a bit hot for future tourism). From its motion, astronomers calculate that it has at least half the mass of Jupiter[footnote], making it clearly a jovian and not a terrestrial-type planet. The Doppler method only allows us to find the minimum mass of a planet. To determine the exact mass using the Doppler shift and Kepler’s laws, we must also have the angle at which the planet’s orbit is oriented to our view—something we don’t have any independent way of knowing in most cases.

Still, if the minimum mass is half of Jupiter's, the actual mass can only be larger than that, and we are sure that we are dealing with a jovian planet. Hot Jupiter.



Artist Greg Bacon painted this impression of a hot, Jupiter-type planet orbiting close to a sunlike star. The artist shows bands on the planet like Jupiter, but we only estimate the mass of most hot, Jupiter-type planets from the Doppler method and don't know what conditions on the planet are like. (credit: ESO)

Since that initial planet discovery, the rate of progress has been breathtaking. Hundreds of giant planets have been discovered using the Doppler technique. Many of these giant planets are orbiting close to their stars—astronomers have called these *hot Jupiters*.

The existence of giant planets so close to their stars was a surprise, and these discoveries have forced us to rethink our ideas about how planetary

systems form. But for now, bear in mind that the Doppler-shift method—which relies on the pull of a planet making its star “wiggle” back and forth around the center of mass—is most effective at finding planets that are both close to their stars and massive. These planets cause the biggest “wiggles” in the motion of their stars and the biggest Doppler shifts in the spectrum. Plus, they will be found sooner, since astronomers like to monitor the star for at least one full orbit (and perhaps more) and hot Jupiters take the shortest time to complete their orbit.

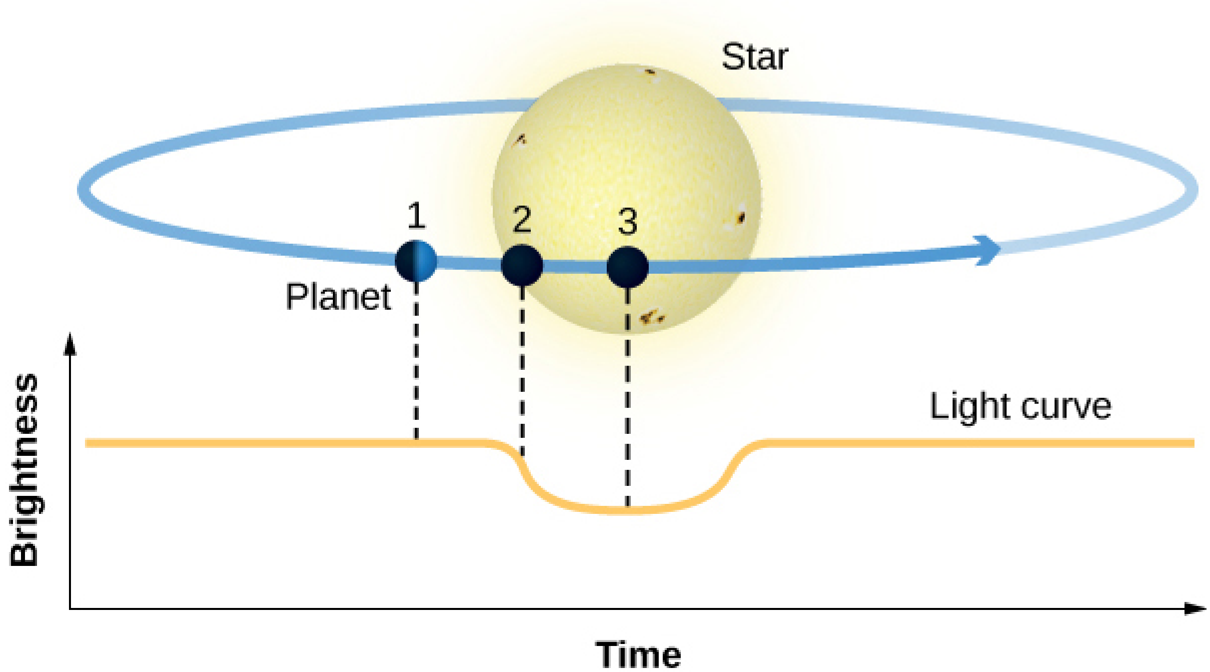
So if such planets exist, we would expect to be finding this type first. Scientists call this a *selection effect*—where our technique of discovery selects certain kinds of objects as “easy finds.” As an example of a selection effect in everyday life, imagine you decide you are ready for a new romantic relationship in your life. To begin with, you only attend social events on campus, all of which require a student ID to get in. Your selection of possible partners will then be limited to students at your college. That may not give you as diverse a group to choose from as you want. In the same way, when we first used the Doppler technique, it selected massive planets close to their stars as the most likely discoveries. As we spend longer times watching target stars and as our ability to measure smaller Doppler shifts improves, this technique can reveal more distant and less massive planets too.

Note:
View a series of animations demonstrating solar system motion and Kepler’s laws, and select animation 1 (Kepler’s laws) from the dropdown playlist. To view an animation demonstrating the radial velocity curve for an exoplanet, select animation 29 (radial velocity curve for an exoplanet) and animation 30 (radial velocity curve for an exoplanet—elliptical orbit) from the dropdown playlist.

Transiting Planets

The second method for indirect detection of exoplanets is based not on the motion of the star but on its brightness. When the orbital plane of the planet is tilted or inclined so that it is viewed edge-on, we will see the planet cross in front of the star once per orbit, causing the star to dim slightly; this event is known as **transit**. [link] shows a sketch of the transit at three time steps: (1) out of transit, (2) the start of transit, and (3) full transit, along with a sketch of the light curve, which shows the drop in the brightness of the host star. The amount of light blocked—the depth of the transit—depends on the area of the planet (its size) compared to the star. If we can determine the size of the star, the transit method tells us the size of the planet.

Planet Transits.



As the planet transits, it blocks out some of the light from the star, causing a temporary dimming in the brightness of the star. The top figure shows three moments during the transit event and the bottom panel shows the corresponding light curve: (1) out of transit, (2) transit ingress, and (3) the full drop in brightness.

The interval between successive transits is the length of the year for that planet, which can be used (again using Kepler's laws) to find its distance

from the star. Larger planets like Jupiter block out more starlight than small earthlike planets, making transits by giant planets easier to detect, even from ground-based observatories. But by going into space, above the distorting effects of Earth's atmosphere, the transit technique has been extended to exoplanets as small as Mars.

Example:

Transit Depth

In a transit, the planet's circular disk blocks the light of the star's circular disk. The area of a circle is πR^2. The amount of light the planet blocks, called the transit depth, is then given by

Equation:

$$\frac{\pi R^2{}_{\text{planet}}}{\pi R^2{}_{\text{star}}} = \frac{R^2{}_{\text{planet}}}{R^2{}_{\text{star}}} = \left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2$$

Now calculate the transit depth for a star the size of the Sun with a gas giant planet the size of Jupiter.

Solution

The radius of Jupiter is 71,400 km, while the radius of the Sun is 695,700 km. Substituting into the equation, we get $\left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2 = \left(\frac{71{,}400 \text{ km}}{695{,}700 \text{ km}}\right)^2 = 0.01$ or 1%, which can easily be detected with the instruments on board the Kepler spacecraft.

Check Your Learning

What is the transit depth for a star half the size of the Sun with a much smaller planet, like the size of Earth?

Note:

Answer:

The radius of Earth is 6371 km. Therefore, $\left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2 = \left(\frac{6371 \text{ km}}{695{,}700/2 \text{ km}}\right)^2 = \left(\frac{6371 \text{ km}}{347{,}850 \text{ km}}\right)^2 = 0.0003$, or significantly less than 1%.

The Doppler method allows us to estimate the mass of a planet. If the same object can be studied by both the Doppler and transit techniques, we can measure both the mass and the size of the exoplanet. This is a powerful combination that can be used to derive the average density (mass/volume) of the planet. In 1999, using measurements from ground-based telescopes, the first transiting planet was detected orbiting the star HD 209458. The planet transits its parent star for about 3 hours every 3.5 days as we view it from Earth. Doppler measurements showed that the planet around HD 209458 has about 70% the mass of Jupiter, but its radius is about 35% larger than Jupiter's. This was the first case where we could determine what an exoplanet was made of—with that mass and radius, HD 209458 must be a gas and liquid world like Jupiter or Saturn.

It is even possible to learn something about the planet's atmosphere. When the planet passes in front of HD 209458, the atoms in the planet's atmosphere absorb starlight. Observations of this absorption were first made at the wavelengths of yellow sodium lines and showed that the atmosphere of the planet contains sodium; now, other elements can be measured as well.

Note:
Try a transit simulator that demonstrates how a planet passing in front of its parent star can lead to the planet's detection. Follow the instructions to run the animation on your computer.

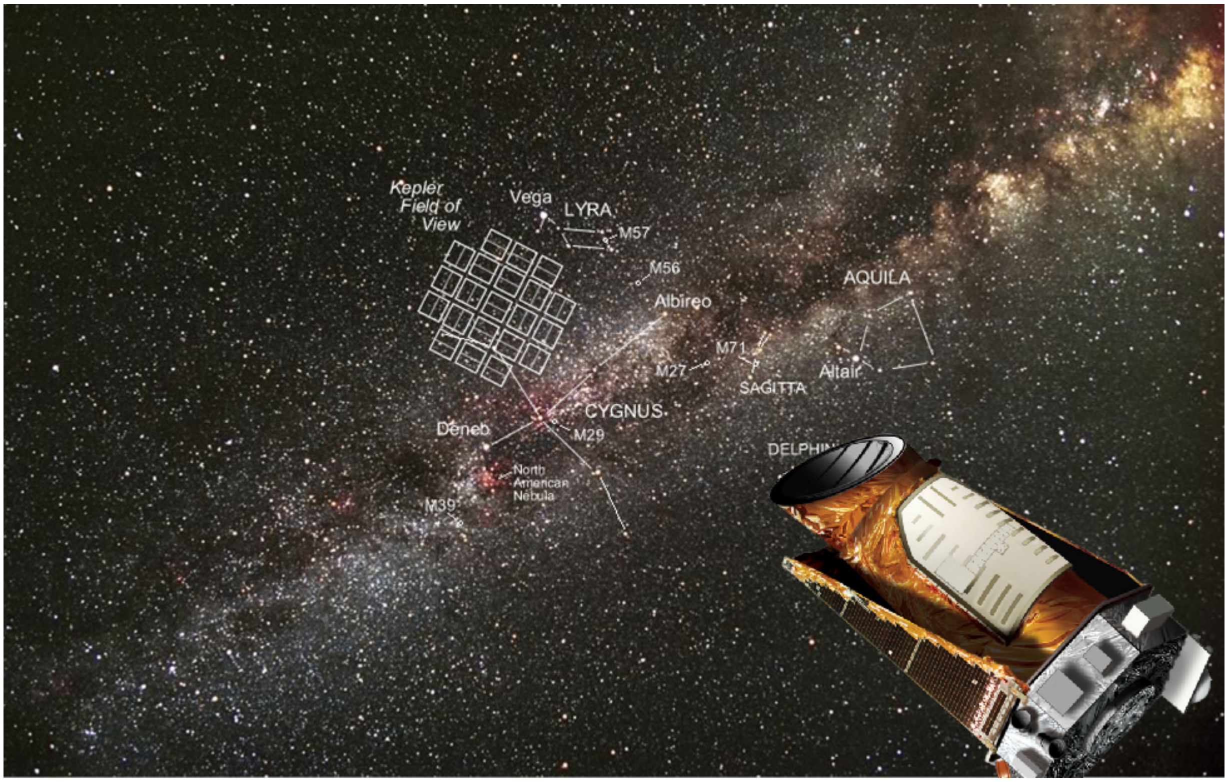
Transiting planets reveal such a wealth of information that the French Space Agency (CNES) and the European Space Agency (ESA) launched the CoRoT space telescope in 2007 to detect transiting exoplanets. CoRoT

discovered 32 transiting exoplanets, including the first transiting planet with a size and density similar to Earth. In 2012, the spacecraft suffered an onboard computer failure, ending the mission. Meanwhile, NASA built a much more powerful transit observatory called Kepler.

In 2009, NASA launched the Kepler space telescope, dedicated to the discovery of transiting exoplanets. This spacecraft stared continuously at more than 150,000 stars in a small patch of sky near the constellation of Cygnus—just above the plane of our Milky Way Galaxy ([link]). Kepler's cameras and ability to measure small changes in brightness very precisely enabled the discovery of thousands of exoplanets, including many multi-planet systems. The spacecraft required three reaction wheels—a type of wheel used to help control slight rotation of the spacecraft—to stabilize the pointing of the telescope and monitor the brightness of the same group of stars over and over again. Kepler was launched with four reaction wheels (one a spare), but by May 2013, two wheels had failed and the telescope could no longer be accurately pointed toward the target area. Kepler had been designed to operate for 4 years, and ironically, the pointing failure occurred exactly 4 years and 1 day after it began observing.

However, this failure did not end the mission. The Kepler telescope continued to observe for two more years, looking for short-period transits in different parts of the sky. A new NASA mission called TESS (Transiting Exoplanet Survey Satellite) is now carrying out a transit survey all over the sky of the nearer (and therefore brighter) stars.

Kepler's Field of View.



The boxes show the region where the Kepler spacecraft cameras took images of over 150,000 stars regularly, to find transiting planets. (credit "field of view": modification of work by NASA/Kepler mission; credit "spacecraft": modification of work by NASA/Kepler mission/Wendy Stenzel)

What do we mean, exactly, by "discovery" of transiting exoplanets? A single transit shows up as a very slight drop in the brightness of the star, lasting several hours. However, astronomers must be on guard against other factors that might produce a false transit, especially when working at the limit of precision of the telescope. We must wait for a second transit of similar depth. But when another transit is observed, we don't initially know whether it might be due to another planet in a different orbit. The "discovery" occurs only when a third transit is found with similar depth and the same spacing in time as the first pair.

Computers normally conduct the analysis, which involves searching for tiny, periodic dips in the light from each star, extending over 4 years of observation. But the Kepler mission also has a program in which non-astronomers—citizen scientists—can examine the data. These dedicated volunteers have found several transits that were missed by the computer analyses, showing that the human eye and brain sometimes recognize unusual events that a computer was not programmed to look for.

Measuring three or four evenly spaced transits is normally enough to "discover" an exoplanet. But in a new field like exoplanet research, we would like to find further independent verification. The strongest confirmation happens when ground-based telescopes are also able to detect a Doppler shift with the same period as the transits. However, this is generally not possible for Earth-size planets. One of the most convincing ways to verify that a dip in brightness is due to a planet is to find more planets orbiting the same star—a *planetary system*. Multi-planet systems also provide alternative ways to estimate the masses of the planets, as we will discuss in the next section.

The selection effects (or biases) in the Kepler data are similar to those in Doppler observations. Large planets are easier to find than small ones, and short-period planets are easier than long-period planets. If we require three transits to establish the presence of a planet, we are of course limited to discovering planets with orbital periods less than one-third of the observing interval. Thus, it was only in its fourth and final year of operation that Kepler was able to find planets with orbits like Earth's that require 1 year to go around their star.

Direct Detection

The best possible evidence for an earthlike planet elsewhere would be an image. After all, "seeing is believing" is a very human prejudice. But imaging a distant planet is a formidable challenge indeed. Suppose, for example, you were a great distance away and wished to detect reflected light from Earth. Earth intercepts and reflects less than one billionth of the Sun's radiation, so its apparent brightness in visible light is less than one billionth that of the Sun. Compounding the challenge of detecting such a

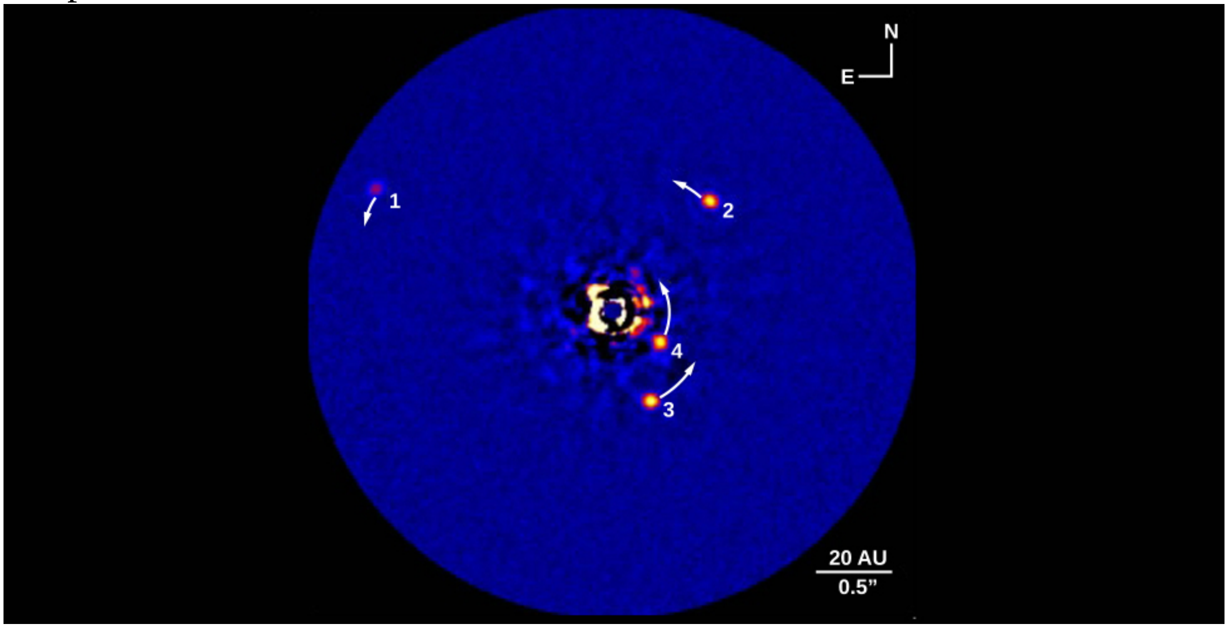
faint speck of light, the planet is swamped by the blaze of radiation from its parent star.

Even today, the best telescope mirrors' optics have slight imperfections that prevent the star's light from coming into focus in a completely sharp point.

Direct imaging works best for young gas giant planets that emit infrared light and reside at large separations from their host stars. Young giant planets emit more infrared light because they have more internal energy, stored from the process of planet formation. Even then, clever techniques must be employed to subtract out the light from the host star. In 2008, three such young planets were discovered orbiting HR 8799, a star in the constellation of Pegasus ([link]). Two years later, a fourth planet was detected closer to the star. Additional planets may reside even closer to HR 8799, but if they exist, they are currently lost in the glare of the star.

Since then, a number of planets around other stars have been found using direct imaging. However, one challenge is to tell whether the objects we are seeing are indeed planets or if they are brown dwarfs (failed stars) in orbit around a star.

Exoplanets around HR 8799.



This image shows Keck telescope observations of four directly imaged planets orbiting HR 8799. A size scale for the system gives the

distance in AU (remember that one astronomical unit is the distance between Earth and the Sun.) (credit: modification of work by Ben Zuckerman)

Direct imaging is an important technique for characterizing an exoplanet. The brightness of the planet can be measured at different wavelengths. These observations provide an estimate for the temperature of the planet's atmosphere; in the case of HR 8799 planet 1, the color suggests the presence of thick clouds. Spectra can also be obtained from the faint light to analyze the atmospheric constituents. A spectrum of HR 8799 planet 1 indicates a hydrogen-rich atmosphere, while the closer planet 4 shows evidence for methane in the atmosphere.

Another way to overcome the blurring effect of Earth's atmosphere is to observe from space. Infrared may be the optimal wavelength range in which to observe because planets get brighter in the infrared while stars like our Sun get fainter, thereby making it easier to detect a planet against the glare of its star. Special optical techniques can be used to suppress the light from the central star and make it easier to see the planet itself. However, even if we go into space, it will be difficult to obtain images of Earth-size planets.

Key Concepts and Summary

Several observational techniques have successfully detected planets orbiting other stars. These techniques fall into two general categories—direct and indirect detection. The Doppler and transit techniques are our most powerful indirect tools for finding exoplanets. Some planets are also being found by direct imaging.

Glossary

exoplanet
: a planet orbiting a star other than our Sun

transit

when one astronomical object moves in front of another

Exoplanets Everywhere: What We Are Learning

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain what we have learned from our discovery of exoplanets
- Identify which kind of exoplanets appear to be the most common in the Galaxy
- Discuss the kinds of planetary systems we are finding around other stars

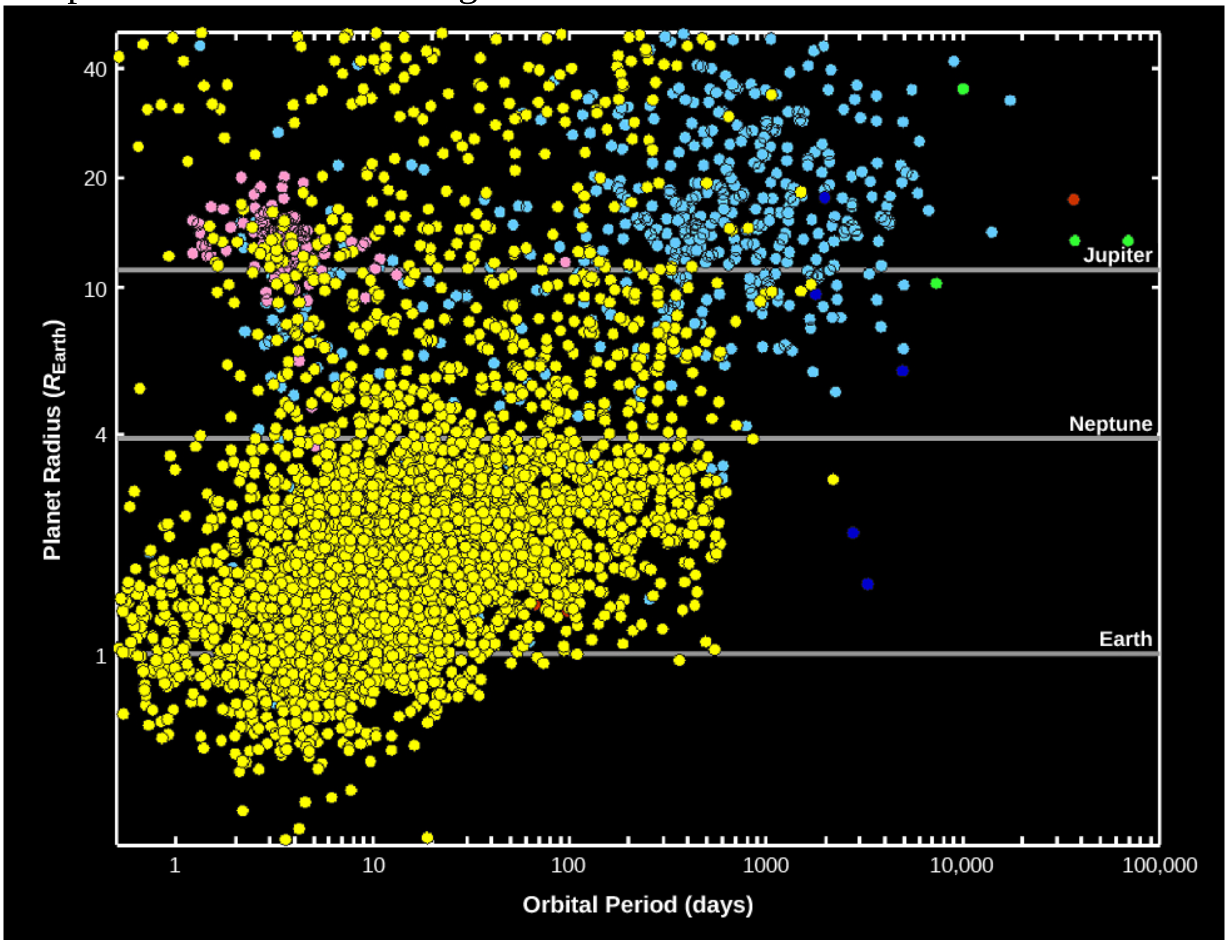
Before the discovery of exoplanets, most astronomers expected that other planetary systems would be much like our own—planets following roughly circular orbits, with the most massive planets several AU from their parent star. Such systems do exist in large numbers, but many exoplanets and planetary systems are very different from those in our solar system. Another surprise is the existence of whole classes of exoplanets that we simply don’t have in our solar system: planets with masses between the mass of Earth and Neptune, and planets that are several times more massive than Jupiter.

Kepler Results

The Kepler telescope has been responsible for the discovery of most exoplanets, especially at smaller sizes, as illustrated in [link], where the Kepler discoveries are plotted in yellow. You can see the wide range of sizes, including planets substantially larger than Jupiter and smaller than Earth. The absence of Kepler-discovered exoplanets with orbital periods

longer than a few hundred days is a consequence of the 4-year lifetime of the mission. (Remember that three evenly spaced transits must be observed to register a discovery.) At the smaller sizes, the absence of planets much smaller than one earth radius is due to the difficulty of detecting transits by very small planets. In effect, the “discovery space” for Kepler was limited to planets with orbital periods less than 400 days and sizes larger than Mars.

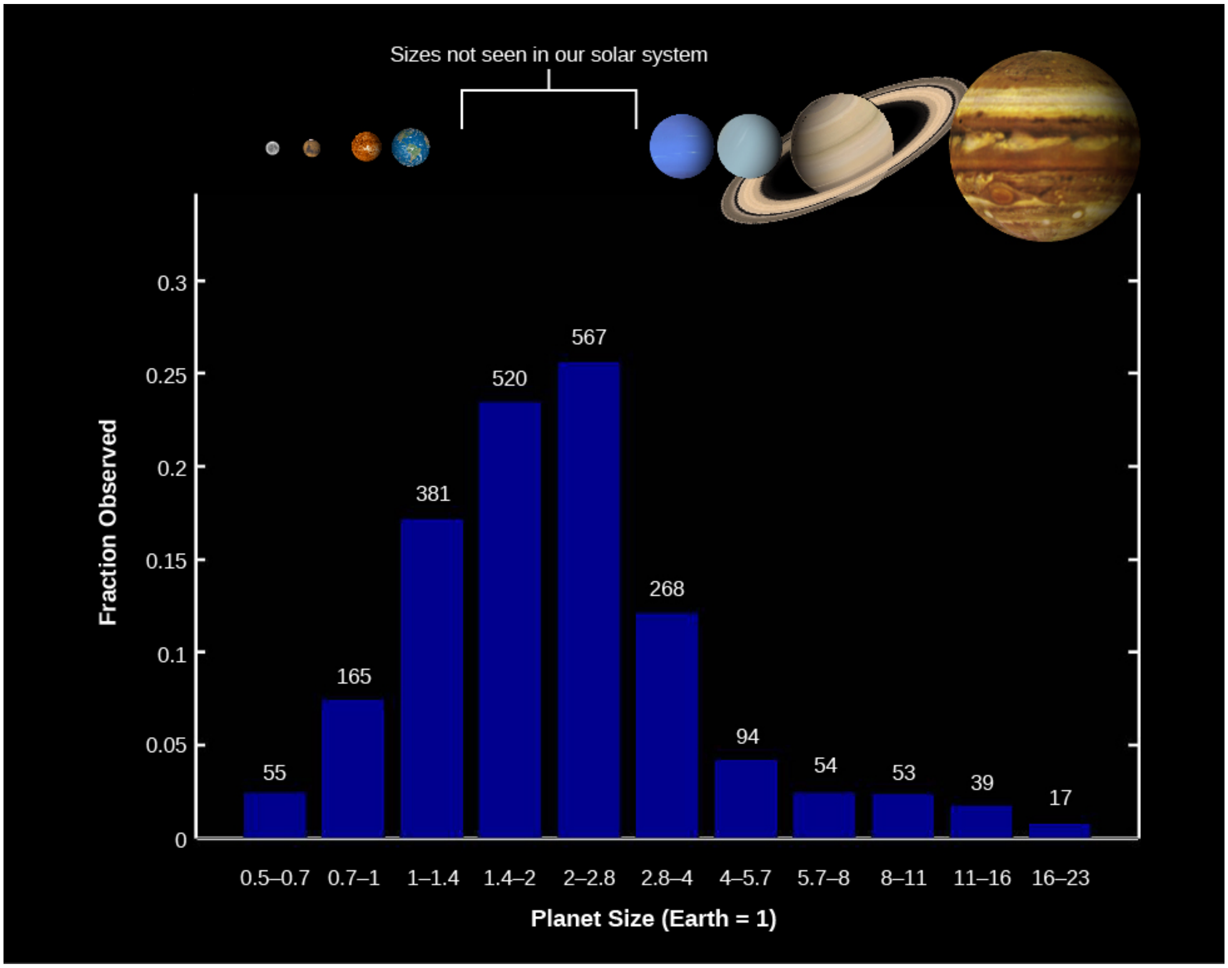
Exoplanet Discoveries through 2015.



The vertical axis shows the radius of each planet compared to Earth. Horizontal lines show the size of Earth, Neptune, and Jupiter. The horizontal axis shows the time each planet takes to make one orbit (and is given in Earth days). Recall that Mercury takes 88 days and Earth takes a little more than 365 days to orbit the Sun. The yellow and red dots show planets discovered by transits, and the blue dots are the discoveries by the radial velocity (Doppler) technique. (credit: modification of work by NASA/Kepler mission)

One of the primary objectives of the Kepler mission was to find out how many stars hosted planets and especially to estimate the frequency of earthlike planets. Although Kepler looked at only a very tiny fraction of the stars in the Galaxy, the sample size was large enough to draw some interesting conclusions. While the observations apply only to the stars observed by Kepler, those stars are reasonably representative, and so astronomers can extrapolate to the entire Galaxy.

[link] shows that the Kepler discoveries include many rocky, Earth-size planets, far more than Jupiter-size gas planets. This immediately tells us that the initial Doppler discovery of many hot Jupiters was a biased sample, in effect, finding the odd planetary systems because they were the easiest to detect. However, there is one huge difference between this observed size distribution and that of planets in our solar system. The most common planets have radii between 1.4 and 2.8 that of Earth, sizes for which we have no examples in the solar system. These have been nicknamed **super-Earths**, while the other large group with sizes between 2.8 and 4 that of Earth are often called **mini-Neptunes**.
Kepler Discoveries.

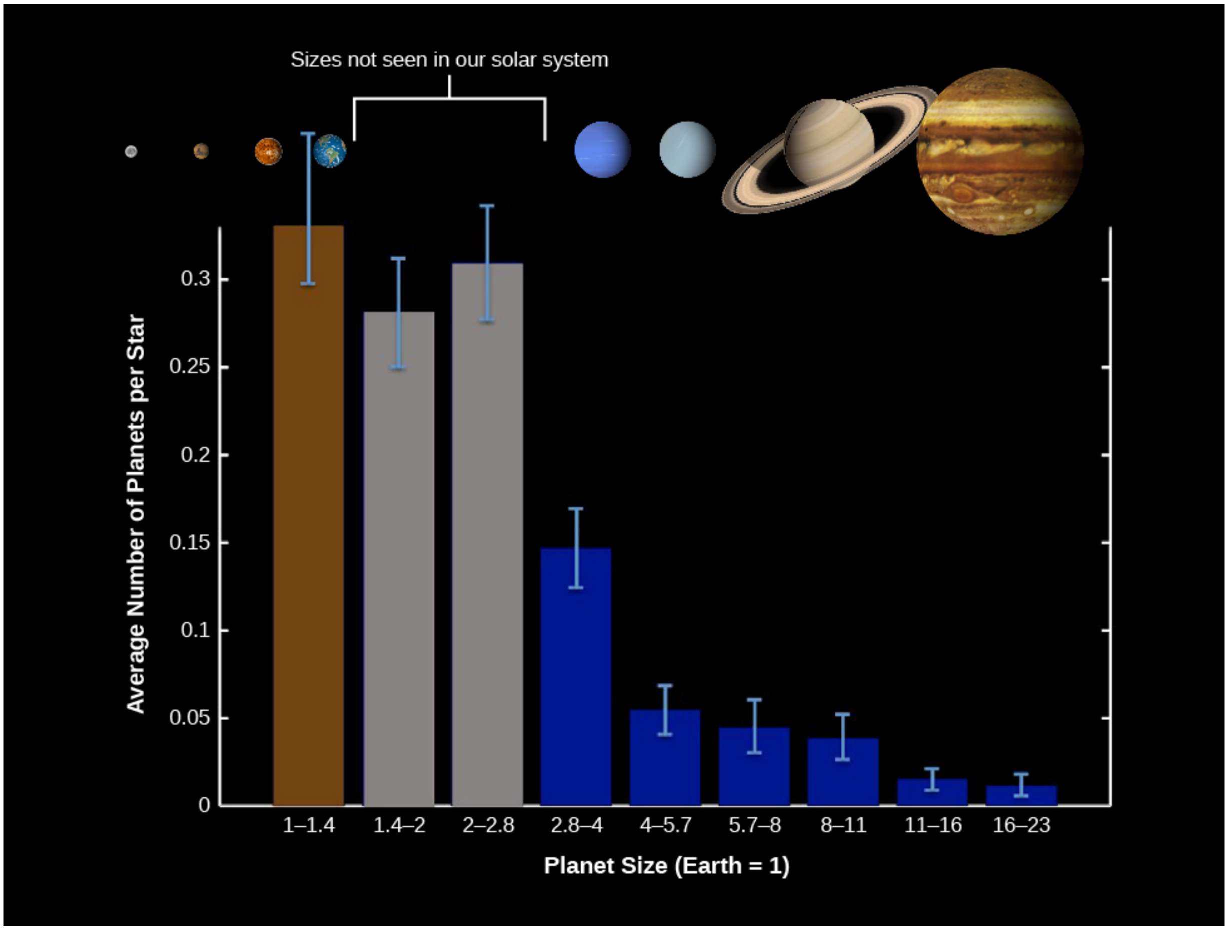


This bar graph shows the number of planets of each size range found among the first 2213 Kepler planet discoveries. Sizes range from half the size of Earth to 20 times that of Earth. On the vertical axis, you can see the fraction that each size range makes up of the total. Note that planets that are between 1.4 and 4 times the size of Earth make up the largest fractions, yet this size range is not represented among the planets in our solar system. (credit: modification of work by NASA/Kepler mission)

What a remarkable discovery it is that the most common types of planets in the Galaxy are completely absent from our solar system and were unknown until Kepler’s survey. However, recall that really small planets were difficult for the Kepler instruments to find. So, to estimate the frequency of

Earth-size exoplanets, we need to correct for this sampling bias. The result is the corrected size distribution shown in [link]. Notice that in this graph, we have also taken the step of showing not the number of Kepler detections but the average number of planets per star for solar-type stars (spectral types F, G, and K).

Size Distribution of Planets for Stars Similar to the Sun.



We show the average number of planets per star in each planet size range. (The average is less than one because some stars will have zero planets of that size range.) This distribution, corrected for biases in the Kepler data, shows that Earth-size planets may actually be the most common type of exoplanets. (credit: modification of work by NASA/Kepler mission)

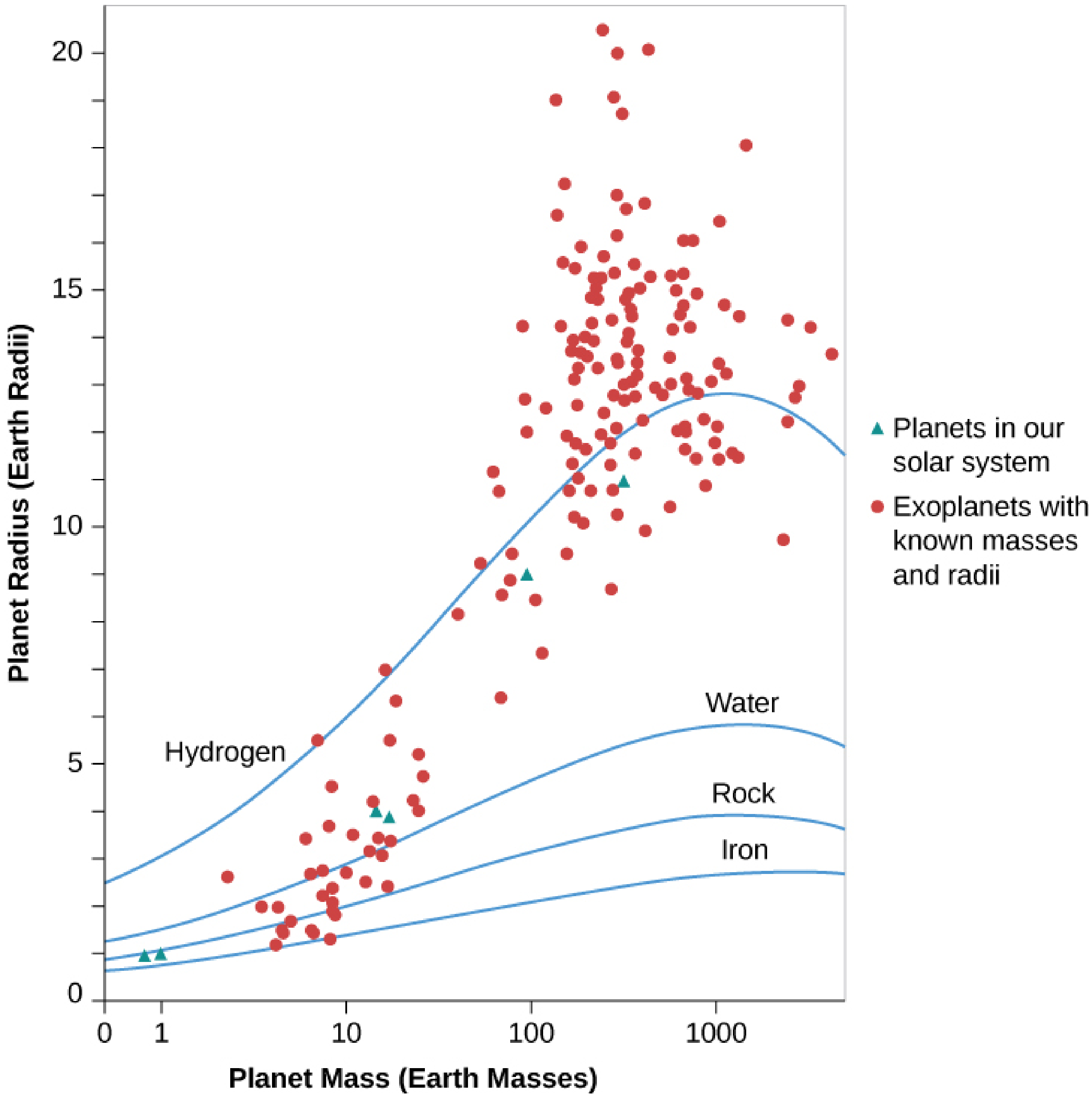
We see that the most common planet sizes of are those with radii from 1 to 3 times that of Earth—what we have called "Earths" and "super-Earths." Each group occurs in about one-third to one-quarter of stars. In other words, if we group these sizes together, we can conclude there is nearly one such planet per star! And remember, this census includes primarily planets with orbital periods less than 2 years. We do not yet know how many undiscovered planets might exist at larger distances from their star.

To estimate the number of Earth-size planets in our Galaxy, we need to remember that there are approximately 100 billion stars of spectral types F, G, and K. Therefore, we estimate that there are about 30 billion Earth-size planets in our Galaxy. If we include the super-Earths too, then there could be one hundred billion in the whole Galaxy. This idea—that planets of roughly Earth's size are so numerous—is surely one of the most important discoveries of modern astronomy.

Planets with Known Densities

For several hundred exoplanets, we have been able to measure both the size of the planet from transit data and its mass from Doppler data, yielding an estimate of its density. Comparing the average density of exoplanets to the density of planets in our solar system helps us understand whether they are rocky or gaseous in nature. This has been particularly important for understanding the structure of the new categories of super-Earths and mini-Neptunes with masses between 3–10 times the mass of Earth. A key observation so far is that planets that are more than 10 times the mass of Earth have substantial gaseous envelopes (like Uranus and Neptune) whereas lower-mass planets are predominately rocky in nature (like the terrestrial planets).

[link] compares all the exoplanets that have both mass and radius measurements. The dependence of the radius on planet mass is also shown for a few illustrative cases—hypothetical planets made of pure iron, rock, water, or hydrogen.
Exoplanets with Known Densities.



Exoplanets with known masses and radii (red circles) are plotted along with solid lines that show the theoretical size of pure iron, rock, water, and hydrogen planets with increasing mass. Masses are given in multiples of Earth's mass. (For comparison, Jupiter contains enough mass to make 320 Earths.) The green triangles indicate planets in our solar system.

At lower masses, notice that as the mass of these hypothetical planets increases, the radius also increases. That makes sense—if you were building a model of a planet out of clay, your toy planet would increase in size as you added more clay. However, for the highest mass planets ($M >$ 1000 M_{Earth}) in [link], notice that the radius stops increasing and the planets with greater mass are actually smaller. This occurs because increasing the mass also increases the gravity of the planet, so that compressible materials (even rock is compressible) will become more tightly packed, shrinking the size of the more massive planet.

In reality, planets are not pure compositions like the hypothetical water or iron planet. Earth is composed of a solid iron core, an outer liquid-iron core, a rocky mantle and crust, and a relatively thin atmospheric layer. Exoplanets are similarly likely to be differentiated into compositional layers. The theoretical lines in [link] are simply guides that suggest a range of possible compositions.

Astronomers who work on the complex modeling of the interiors of rocky planets make the simplifying assumption that the planet consists of two or three layers. This is not perfect, but it is a reasonable approximation and another good example of how science works. Often, the first step in understanding something new is to narrow down the range of possibilities. This sets the stage for refining and deepening our knowledge. In [link], the two green triangles with roughly 1 M_{Earth} and 1 R_{Earth} represent Venus and Earth. Notice that these planets fall between the models for a pure iron and a pure rock planet, consistent with what we would expect for the known mixed-chemical composition of Venus and Earth.

In the case of gaseous planets, the situation is more complex. Hydrogen is the lightest element in the periodic table, yet many of the detected exoplanets in [link] with masses greater than 100 M_{Earth} have radii that suggest they are lower in density than a pure hydrogen planet. Hydrogen is the lightest element, so what is happening here? Why do some gas giant planets have inflated radii that are larger than the fictitious pure hydrogen planet? Many of these planets reside in short-period orbits close to the host star where they intercept a significant amount of radiated energy. If this

energy is trapped deep in the planet atmosphere, it can cause the planet to expand.

Planets that orbit close to their host stars in slightly eccentric orbits have another source of energy: the star will raise tides in these planets that tend to circularize the orbits. This process also results in tidal dissipation of energy that can inflate the atmosphere. It would be interesting to measure the size of gas giant planets in wider orbits where the planets should be cooler—the expectation is that unless they are very young, these cooler gas giant exoplanets (sometimes called “cold Jupiters”) should not be inflated. But we don’t yet have data on these more distant exoplanets.

Exoplanetary Systems

As we search for exoplanets, we don’t expect to find only one planet per star. Our solar system has eight major planets, half a dozen dwarf planets, and millions of smaller objects orbiting the Sun. The evidence we have of planetary systems in formation also suggest that they are likely to produce multi-planet systems.

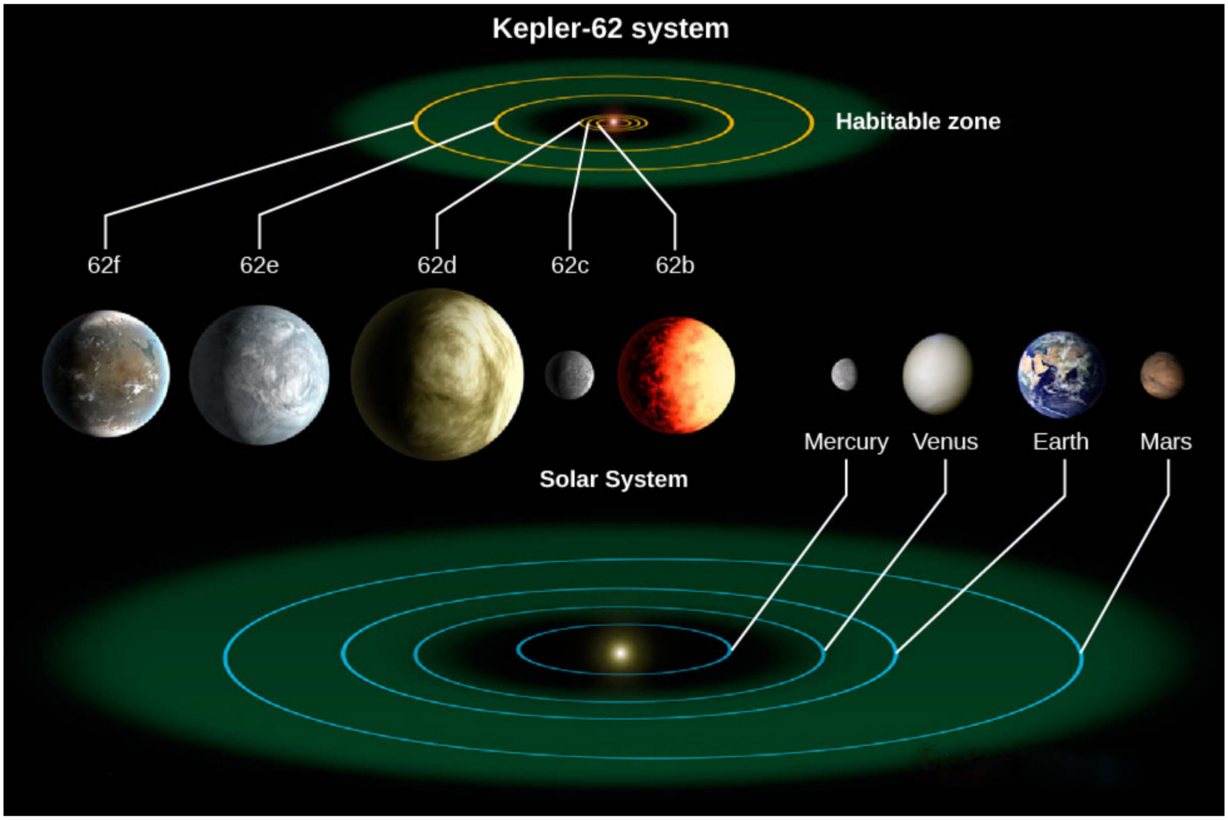
The first planetary system was found around the star Upsilon Andromedae in 1999 using the Doppler method, and many others have been found since then (almost 700 as of early 2020). If such exoplanetary system are common, let’s consider which systems we expect to find in the Kepler transit data.

A planet will transit its star only if Earth lies in the plane of the planet’s orbit. If the planets in other systems do not have orbits in the same plane, we are unlikely to see multiple transiting objects. Also, as we have noted before, Kepler was sensitive only to planets with orbital periods less than about 4 years. What we expect from Kepler data, then, is evidence of coplanar planetary systems confined to what would be the realm of the terrestrial planets in our solar system.

By 2020, astronomers gathered data on nearly 700 such exoplanet systems. Many have only two known planets, but a few have as many as five, and one has eight (the same number of planets as our own solar system). For the

most part, these are very compact systems with most of their planets closer to their star than Mercury is to the Sun. The figure below shows one of the largest exoplanet systems: that of the star called Kepler-62 ([link]). Our solar system is shown to the same scale, for comparison (note that the Kepler-62 planets are drawn with artistic license; we have no detailed images of any exoplanets).

Exoplanet System Kepler-62, with the Solar System Shown to the Same Scale.



The green areas are the “habitable zones,” the range of distance from the star where surface temperatures are likely to be consistent with liquid water. (credit: modification of work by NASA/Ames/JPL-Caltech)

All but one of the planets in the K-62 system are larger than Earth. These are super-Earths, and one of them (62d) is in the size range of a mini-Neptune, where it is likely to be largely gaseous. The smallest planet in this

system is about the size of Mars. The three inner planets orbit very close to their star, and only the outer two have orbits larger than Mercury in our system. The green areas represent each star's "habitable zone," which is the distance from the star where we calculate that surface temperatures would be consistent with liquid water. The Kepler-62 habitable zone is much smaller than that of the Sun because the star is intrinsically fainter.

With closely spaced systems like this, the planets can interact gravitationally with each other. The result is that the observed transits occur a few minutes earlier or later than would be predicted from simple orbits. These gravitational interactions have allowed the Kepler scientists to calculate masses for the planets, providing another way to learn about exoplanets.

Kepler has discovered some interesting and unusual planetary systems. For example, most astronomers expected planets to be limited to single stars. But we have found planets orbiting close double stars, so that the planet would see two suns in its sky, like those of the fictional planet Tatooine in the *Star Wars* films. At the opposite extreme, planets can orbit one star of a wide, double-star system without major interference from the second star.

Key Concepts and Summary

Although the Kepler mission is finding thousands of new exoplanets, these are limited to orbital periods of less than 400 days and sizes larger than Mars. Still, we can use the Kepler discoveries to extrapolate the distribution of planets in our Galaxy. The data so far imply that planets like Earth are the most common type of planet, and that there may be 100 billion Earth-size planets around Sun-like stars in the Galaxy. Almost 700 planetary systems have been discovered around other stars. In many of them, planets are arranged differently than in our solar system.

Glossary

super-Earth
: a planet larger than Earth, generally between 1.4 and 2.8 times the size of our planet

mini-Neptune

a planet that is intermediate between the largest terrestrial planet in our solar system (Earth) and the smallest jovian planet (Neptune); generally, mini-Neptunes have sizes between 2.8 and 4 times Earth’s size

New Perspectives on Planet Formation

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how exoplanet discoveries have revised our understanding of planet formation
- Discuss how planetary systems quite different from our solar system might have come about

Traditionally, astronomers have assumed that the planets in our solar system formed at about their current distances from the Sun and have remained there ever since. The first step in the formation of a giant planet is to build up a solid core, which happens when planetesimals collide and stick. Eventually, this core becomes massive enough to begin sweeping up gaseous material in the disk, thereby building the gas giants Jupiter and Saturn.

How to Make a Hot Jupiter

The traditional model for the formation of planets works only if the giant planets are formed far from the central star (about 5–10 AU), where the disk is cold enough to have a fairly high density of solid matter. It cannot explain the hot Jupiters, which are located very close to their stars where any rocky raw material would be completely vaporized. It also cannot explain the elliptical orbits we observe for some exoplanets because the orbit of a protoplanet, whatever its initial shape, will quickly become

circular through interactions with the surrounding disk of material and will remain that way as the planet grows by sweeping up additional matter.

So we have two options: either we find a new model for forming planets close to the searing heat of the parent star, or we find a way to change the orbits of planets so that cold Jupiters can travel inward *after* they form. Most research now supports the latter explanation.

Calculations show that if a planet forms while a substantial amount of gas remains in the disk, then some of the planet's orbital angular momentum can be transferred to the disk. As it loses momentum (through a process that reminds us of the effects of friction), the planet will spiral inward. This process can transport giant planets, initially formed in cold regions of the disk, closer to the central star—thereby producing hot Jupiters. Gravitational interactions between planets in the chaotic early solar system can also cause planets to slingshot inward from large distances. But for this to work, the other planet has to carry away the angular momentum and move to a more distant orbit.

In some cases, we can use the combination of transit plus Doppler measurements to determine whether the planets orbit in the same plane and in the same direction as the star. For the first few cases, things seemed to work just as we anticipated: like the solar system, the gas giant planets orbited in their star's equatorial plane and in the same direction as the spinning star.

Then, some startling discoveries were made of gas giant planets that orbited at right angles or even in the opposite sense as the spin of the star. How could this happen? Again, there must have been interactions between planets. It's possible that before the system settled down, two planets came close together, so that one was kicked into an unusual orbit. Or perhaps a passing star perturbed the system after the planets were newly formed.

Forming Planetary Systems

When the Milky Way Galaxy was young, the stars that formed did not contain many heavy elements like iron. Several generations of star

formation and star death were required to enrich the interstellar medium for subsequent generations of stars. Since planets seem to form “inside out,” starting with the accretion of the materials that can make the rocky cores with which planets start, astronomers wondered when in the history of the Galaxy, planet formation would turn on.

The star Kepler-444 has shed some light on this question. This is a tightly packed system of five planets—the smallest comparable in size to Mercury and the largest similar in size to Venus. All five planets were detected with the Kepler spacecraft as they transited their parent star. All five planets orbit their host star in less than the time it takes Mercury to complete one orbit about the Sun. Remarkably, the host star Kepler-444 is more than 11 billion years old and formed when the Milky Way was only 2 billion years old. So the heavier elements needed to make rocky planets must have already been available then. This ancient planetary system sets the clock on the beginning of rocky planet formation to be relatively soon after the formation of our Galaxy.

Kepler data demonstrate that while rocky planets inside Mercury’s orbit are missing from our solar system, they are common around other stars, like Kepler-444. When the first systems packed with close-in rocky planets were discovered, we wondered why they were so different from our solar system. When many such systems were discovered, we began to wonder if it was our solar system that was different. This led to speculation that additional rocky planets might once have existed close to the Sun in our solar system.

There is some evidence from the motions in the outer solar system that Jupiter may have migrated inward long ago. If correct, then gravitational perturbations from Jupiter could have dislodged the orbits of close-in rocky planets, causing them to fall into the Sun. Consistent with this picture, astronomers now think that Uranus and Neptune probably did not form at their present distances from the Sun but rather closer to where Jupiter and Saturn are now. The reason for this idea is that density in the disk of matter surrounding the Sun at the time the planets formed was so low outside the orbit of Saturn that it would take several billion years to build up Uranus and Neptune. Yet we saw earlier in the chapter that the disks around protostars survive only a few million years.

Therefore, scientists have developed computer models demonstrating that Uranus and Neptune could have formed near the current locations of Jupiter and Saturn, and then been kicked out to larger distances through gravitational interactions with their neighbors. All these wonderful new observations illustrate how dangerous it can be to draw conclusions about a phenomenon in science (in this case, how planetary systems form and arrange themselves) when you are only working with a single example.

Exoplanets have given rise to a new picture of planetary system formation —one that is much more chaotic than we originally thought. If we think of the planets as being like skaters in a rink, our original model (with only our own solar system as a guide) assumed that the planets behaved like polite skaters, all obeying the rules of the rink and all moving in nearly the same direction, following roughly circular paths. The new picture corresponds more to a roller derby, where the skaters crash into one another, change directions, and sometimes are thrown entirely out of the rink.

Habitable Exoplanets

While thousands of exoplanets have been discovered in the past two decades, every observational technique has fallen short of finding more than a few candidates that resemble Earth ([link]). Astronomers are not sure exactly what properties would define another Earth. Do we need to find a planet that is *exactly* the same size and mass as Earth? That may be difficult and may not be important from the perspective of habitability. After all, we have no reason to think that life could not have arisen on Earth if our planet had been a little bit smaller or larger. And, remember that how habitable a planet is depends on both its distance from its star and the nature of its atmosphere. The greenhouse effect can make some planets warmer (as it did for Venus and is doing more and more for Earth).

Many Earthlike Planets.

This painting, commissioned by NASA, conveys the idea that there may be many planets resembling Earth out there as our methods for finding them improve. (credit: NASA/JPL-Caltech/R. Hurt (SSC-Caltech))

We can ask other questions to which we don’t yet know the answers. Does this “twin” of Earth need to orbit a solar-type star, or can we consider as candidates the numerous exoplanets orbiting K- and M-class stars? (In the summer of 2016, astronomers reported the discovery of a planet with at least 1.3 times the mass of Earth around the nearest star, Proxima Centauri, which is spectral type M and located 4.2 light years from us.) We have a special interest in finding planets that could support life like ours, in which case, we need to find exoplanets within their star’s habitable zone, where

surface temperatures are consistent with liquid water on the surface. This is probably the most important characteristic defining an Earth-analog exoplanet.

The search for potentially habitable worlds is one of the prime drivers for exoplanet research in the next decade. Astronomers are beginning to develop realistic plans for new instruments that can even look for signs of life on distant worlds (examining their atmospheres for gases associated with life, for example). If we require telescopes in space to find such worlds, we need to recognize that years are required to plan, build, and launch such space observatories. The discovery of exoplanets and the knowledge that most stars have planetary systems are transforming our thinking about life beyond Earth. We are closer than ever to knowing whether habitable (and inhabited) planets are common. This work lends a new spirit of optimism to the search for life elsewhere, a subject to which we will return in Life in the Universe.

Note:
Check out the habitability of various stars and planets by trying out the interactive Circumstellar Habitable Zone Simulator and select a star system to investigate.

Key Concepts and Summary

The ensemble of exoplanets is incredibly diverse and has led to a revision in our understanding of planet formation that includes the possibility of vigorous, chaotic interactions, with planet migration and scattering. It is possible that the solar system is unusual (and not representative) in how its planets are arranged. Many systems seem to have rocky planets farther inward than we do, for example, and some even have “hot Jupiters” very close to their star. Ambitious space experiments should make it possible to image earthlike planets outside the solar system and even to obtain information about their habitability as we search for life elsewhere.

For Further Exploration

Articles

Star Formation

Blaes, O. "A Universe of Disks." *Scientific American* (October 2004): 48. On accretion disks and jets around young stars and black holes.

Croswell, K. "The Dust Belt Next Door [Tau Ceti]." *Scientific American* (January 2015): 24. Short intro to recent observations of planets and a wide dust belt.

Frank, A. "Starmaker: The New Story of Stellar Birth." *Astronomy* (July 1996): 52.

Jayawardhana, R. "Spying on Stellar Nurseries." *Astronomy* (November 1998): 62. On protoplanetary disks.

O'Dell, C. R. "Exploring the Orion Nebula." *Sky & Telescope* (December 1994): 20. Good review with Hubble results.

Ray, T. "Fountains of Youth: Early Days in the Life of a Star." *Scientific American* (August 2000): 42. On outflows from young stars.

Young, E. "Cloudy with a Chance of Stars." *Scientific American* (February 2010): 34. On how clouds of interstellar matter turn into star systems.

Young, Monica "Making Massive Stars." *Sky & Telescope* (October 2015): 24. Models and observations on how the most massive stars form.

Exoplanets

Billings, L. "In Search of Alien Jupiters." *Scientific American* (August 2015): 40–47. The race to image jovian planets with current instruments

and why a direct image of a terrestrial planet is still in the future.

Heller, R. “Better Than Earth.” *Scientific American* (January 2015): 32–39. What kinds of planets may be habitable; super-Earths and jovian planet moons should also be considered.

Laughlin, G. “How Worlds Get Out of Whack.” *Sky & Telescope* (May 2013): 26. On how planets can migrate from the places they form in a star system.

Marcy, G. “The New Search for Distant Planets.” *Astronomy* (October 2006): 30. Fine brief overview. (The same issue has a dramatic fold-out visual atlas of extrasolar planets, from that era.)

Redd, N. “Why Haven’t We Found Another Earth?” *Astronomy* (February 2016): 25. Looking for terrestrial planets in the habitable zone with evidence of life.

Seager, S. “Exoplanets Everywhere.” *Sky & Telescope* (August 2013): 18. An excellent discussion of some of the frequently asked questions about the nature and arrangement of planets out there.

Seager, S. “The Hunt for Super-Earths.” *Sky & Telescope* (October 2010): 30. The search for planets that are up to 10 times the mass of Earth and what they can teach us.

Villard, R. “Hunting for Earthlike Planets.” *Astronomy* (April 2011): 28. How we expect to find and characterize super-Earth (planets somewhat bigger than ours) using new instruments and techniques that could show us what their atmospheres are made of.

Websites

Exoplanet Exploration: http://planetquest.jpl.nasa.gov/. PlanetQuest (from the Navigator Program at the Jet Propulsion Lab) is probably the best site for students and beginners, with introductory materials and nice illustrations; it focuses mostly on NASA work and missions.

Exoplanets: http://www.planetary.org/exoplanets/. Planetary Society's exoplanets pages with a dynamic catalog of planets found and good explanations.

Exoplanets: The Search for Planets beyond Our Solar System: http://www.iop.org/publications/iop/2010/page_42551.html. From the British Institute of Physics in 2010.

Extrasolar Planets Encyclopedia: http://exoplanet.eu/. Maintained by Jean Schneider of the Paris Observatory, has the largest catalog of planet discoveries and useful background material (some of it more technical).

Formation of Stars: https://www.spacetelescope.org/science/formation_of_stars/. Star Formation page from the Hubble Space Telescope, with links to images and information.

Kepler Mission: http://kepler.nasa.gov/. The public website for the remarkable telescope in space that is searching planets using the transit technique and is our best hope for finding earthlike planets.

Proxima Centauri Planet Discovery: http://www.eso.org/public/news/eso1629/.

Apps

Exoplanet: http://itunes.apple.com/us/app/exoplanet/id327702034?mt=8. Allows you to browse through a regularly updated visual catalog of exoplanets that have been found so far.

Journey to the Exoplanets: http://itunes.apple.com/us/app/journey-to-the-exoplanets/id463532472?mt=8. Produced by the staff of *Scientific American*, with input from scientists and space artists; gives background information and visual tours of the nearer star systems with planets.

Videos

A Star Is Born: http://www.discovery.com/tv-shows/other-shows/videos/how-the-universe-works-a-star-is-born/. Discovery Channel video with astronomer Michelle Thaller (2:25).

Are We Alone: An Evening Dialogue with the Kepler Mission Leaders: http://www.youtube.com/watch?v=O7ItAXfl0Lw. A non-technical panel discussion on Kepler results and ideas about planet formation with Bill Borucki, Natalie Batalha, and Gibor Basri (moderated by Andrew Fraknoi) at the University of California, Berkeley (2:07:01).

Finding the Next Earth: The Latest Results from Kepler: https://www.youtube.com/watch?v=ZbijeR_AALo. Natalie Batalha (San Jose State University & NASA Ames) public talk in the Silicon Valley Astronomy Lecture Series (1:28:38).

From Hot Jupiters to Habitable Worlds: https://vimeo.com/37696087 (Part 1) and https://vimeo.com/37700700 (Part 2). Debra Fischer (Yale University) public talk in Hawaii sponsored by the Keck Observatory (15:20 Part 1, 21:32 Part 2).

Search for Habitable Exoplanets: http://www.youtube.com/watch?v=RLWb_T9yaDU. Sara Seeger (MIT) public talk at the SETI Institute, with Kepler results (1:10:35).

Strange Planetary Vistas: http://www.youtube.com/watch?v=_8ww9eLRSCg. Josh Carter (CfA) public talk at Harvard's Center for Astrophysics with a friendly introduction to exoplanets for non-specialists (46:35).

Collaborative Group Activities

A. Your group is a subcommittee of scientists examining whether any of the "hot Jupiters" (giant planets closer to their stars than Mercury is to the Sun) could have life on or near them. Can you come up with places on, in, or near such planets where life could develop or where some forms of life might survive?

B. A wealthy couple (who are alumni of your college or university and love babies) leaves the astronomy program several million dollars in their will, to spend in the best way possible to search for “infant stars in our section of the Galaxy.” Your group has been assigned the task of advising the dean on how best to spend the money. What kind of instruments and search programs would you recommend, and why?
C. Some people consider the discovery of any planets (even hot Jupiters) around other stars one of the most important events in the history of astronomical research. Some astronomers have been surprised that the public is not more excited about the planet discoveries. One reason that has been suggested for this lack of public surprise and excitement is that science fiction stories have long prepared us for there being planets around other stars. (The Starship Enterprise on the 1960s *Star Trek* TV series found some in just about every weekly episode.) What does your group think? Did you know about the discovery of planets around other stars before taking this course? Do you consider it exciting? Were you surprised to hear about it? Are science fiction movies and books good or bad tools for astronomy education in general, do you think?
D. What if future space instruments reveal an earthlike exoplanet with significant amounts of oxygen and methane in its atmosphere? Suppose the planet and its star are 50 light-years away. What does your group suggest astronomers do next? How much effort and money would you recommend be put into finding out more about this planet and why?
E. Discuss with your group the following question: which is easier to find orbiting a star with instruments we have today: a jovian planet or a proto-planetary disk? Make a list of arguments for each side of this question.
F. (This activity should be done when your group has access to the internet.) Go to the page which indexes all the publicly released Hubble Space Telescope images by subject: http://hubblesite.org/newscenter/archive/browse/image/. Under “Star,” go to “Protoplanetary Disk” and find a system—not mentioned in this chapter—that your group likes, and prepare a short report to the class about why you find it interesting. Then, under “Nebula,” go to “Emission” and find a region of star formation not mentioned in this

chapter, and prepare a short report to the class about what you find interesting about it.

G. There is a “citizen science” website called Planet Hunters (http://www.planethunters.org/) where you can participate in identifying exoplanets from the data that Kepler provided. Your group should access the site, work together to use it, and classify two light curves. Report back to the class on what you have done.
H. Yuri Milner, a Russian-American billionaire, recently pledged $100 million to develop the technology to send many miniaturized probes to a star in the Alpha Centauri triple star system (which includes Proxima Centauri, the nearest star to us, now known to have at least one planet.) Each tiny probe will be propelled by powerful lasers at 20% the speed of light, in the hope that one or more might arrive safely and be able to send back information about what it’s like there. Your group should search online for more information about this project (called “Breakthrough: Starshot”) and discuss your reactions to this project. Give specific reasons for your arguments.

Review Questions

Exercise:

Problem:

Give several reasons the Orion molecular cloud is such a useful “laboratory” for studying the stages of star formation.

Exercise:

Problem:

Why is star formation more likely to occur in cold molecular clouds than in regions where the temperature of the interstellar medium is several hundred thousand degrees?

Exercise:

Problem:

Why have we learned a lot about star formation since the invention of detectors sensitive to infrared radiation?

Exercise:

Problem:

Describe what happens when a star forms. Begin with a dense core of material in a molecular cloud and trace the evolution up to the time the newly formed star reaches the main sequence.

Exercise:

Problem:

Describe how the T Tauri star stage in the life of a low-mass star can lead to the formation of a Herbig-Haro (H-H) object.

Exercise:

Problem:

Look at the four stages shown in [link]. In which stage(s) can we see the star in visible light? In infrared radiation?

Exercise:

Problem:

The evolutionary track for a star of 1 solar mass remains nearly vertical in the H–R diagram for a while (see [link]). How is its luminosity changing during this time? Its temperature? Its radius?

Exercise:

Problem:

Two protostars, one 10 times the mass of the Sun and one half the mass of the Sun are born at the same time in a molecular cloud. Which one will be first to reach the main sequence stage, where it is stable and getting energy from fusion?

Exercise:

Problem:

Compare the scale (size) of a typical dusty disk around a forming star with the scale of our solar system.

Exercise:

Problem:

Why is it so hard to see planets around other stars and so easy to see them around our own?

Exercise:

Problem:

Why did it take astronomers until 1995 to discover the first exoplanet orbiting another star like the Sun?

Exercise:

Problem:

Which types of planets are most easily detected by Doppler measurements? By transits?

Exercise:

Problem:

List three ways in which the exoplanets we have detected have been found to be different from planets in our solar system.

Exercise:

Problem:

List any similarities between discovered exoplanets and planets in our solar system.

Exercise:

Problem:

What revisions to the theory of planet formation have astronomers had to make as a result of the discovery of exoplanets?

Exercise:

Problem:

Why are young Jupiters easier to see with direct imaging than old Jupiters?

Thought Questions

Exercise:

Problem:

A friend of yours who did not do well in her astronomy class tells you that she believes all stars are old and none could possibly be born today. What arguments would you use to persuade her that stars are being born somewhere in the Galaxy during your lifetime?

Exercise:

Problem:

Observations suggest that it takes more than 3 million years for the dust to begin clearing out of the inner regions of the disks surrounding protostars. Suppose this is the minimum time required to form a planet. Would you expect to find a planet around a 10-M_{Sun} star? (Refer to [link].)

Exercise:

Problem:

Suppose you wanted to observe a planet around another star with direct imaging. Would you try to observe in visible light or in the infrared? Why? Would the planet be easier to see if it were at 1 AU or 5 AU from its star?

Exercise:

Problem:

Why were giant planets close to their stars the first ones to be discovered? Why has the same technique not been used yet to discover giant planets at the distance of Saturn?

Exercise:

Problem:

Exoplanets in eccentric orbits experience large temperature swings during their orbits. Suppose you had to plan for a mission to such a planet. Based on Kepler's second law, does the planet spend more time closer or farther from the star? Explain.

Figuring for Yourself

Exercise:

Problem:

When astronomers found the first giant planets with orbits of only a few days, they did not know whether those planets were gaseous and liquid like Jupiter or rocky like Mercury. The observations of HD 209458 settled this question because observations of the transit of the star by this planet made it possible to determine the radius of the planet. Use the data given in the text to estimate the density of this planet, and then use that information to explain why it must be a gas giant.

Exercise:

Problem:

An exoplanetary system has two known planets. Planet X orbits in 290 days and Planet Y orbits in 145 days. Which planet is closest to its host star? If the star has the same mass as the Sun, what is the semi-major axis of the orbits for Planets X and Y?

Exercise:

Problem:

Kepler's third law says that the orbital period (in years) is proportional to the square root of the cube of the mean distance (in AU) from the Sun ($P \propto a^{1.5}$). For mean distances from 0.1 to 32 AU, calculate and plot a curve showing the expected Keplerian period. For each planet in our solar system, look up the mean distance from the Sun in AU and the orbital period in years and overplot these data on the theoretical Keplerian curve.

Exercise:

Problem:

Calculate the transit depth for an M dwarf star that is 0.3 times the radius of the Sun with a gas giant planet the size of Jupiter.

Exercise:

Problem:

If a transit depth of 0.00001 can be detected with the Kepler spacecraft, what is the smallest planet that could be detected around a 0.3 R_{sun} M dwarf star?

Exercise:

Problem:

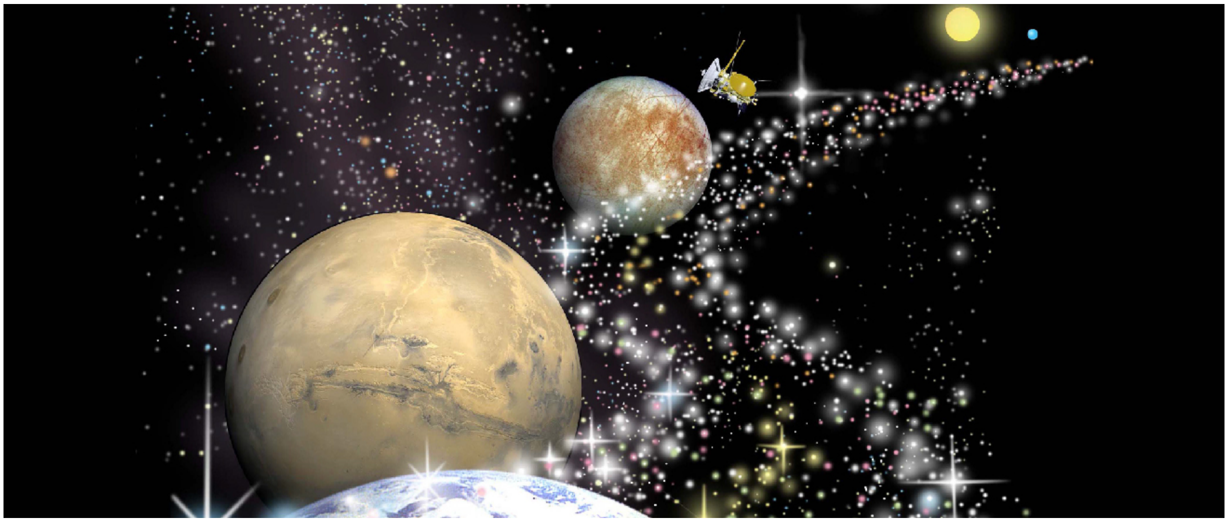
What fraction of gas giant planets seems to have inflated radii?

Thinking Ahead
class="introduction"
Astrobiology: The Road to Life in the Universe.

In this fanciful montage produced by a NASA artist, we see one roadmap for discovering life in the universe. Learning more about the origin, evolution, and properties of life on Earth aids us in searching for evidence of life beyond our planet. Our neighbor world, Mars, had warmer, wetter conditions billions of years ago that might

have helped life there begin. Farther out, Jupiter’s moon Europa represents the icy moons of the outer solar system. Beneath their shells of solid ice may lie vast oceans of liquid water that could support biology. Beyond our solar system are stars that host their own planets, some of which might be similar to Earth in the ability to support liquid water—and a thriving biosphere—

at the planet’s surface. Research is pushing actively in all these directions with the goal of proving a scientific answer to the question, “Are we alone?” (credit: modification of work by NASA)



As we have learned more about the universe, we have naturally wondered whether there might be other forms of life out there. The ancient question,

“Are we alone in the universe?” connects us to generations of humans before us. While in the past, this question was in the realm of philosophy or science fiction, today we have the means to seek an answer through scientific inquiry. In this chapter, we will consider how life began on Earth, whether the same processes could have led to life on other worlds, and how we might seek evidence of life elsewhere. This is the science of astrobiology.

The search for life on other planets is not the same as the search for *intelligent* life, which (if it exists) is surely much rarer. Learning more about the origin, evolution, and properties of life on Earth aids us in searching for evidence of all kinds of life beyond that on our planet.

The Cosmic Context for Life

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the chemical and environmental conditions that make Earth hospitable to life
- Discuss the assumption underlying the Copernican principle and outline its implications for modern-day astronomers
- Understand the questions underlying the Fermi paradox

We saw that the universe was born in the Big Bang about 14 billion years ago. After the initial hot, dense fireball of creation cooled sufficiently for atoms to exist, all matter consisted of hydrogen and helium (with a very small amount of lithium). As the universe aged, processes within stars created the other elements, including those that make up Earth (such as iron, silicon, magnesium, and oxygen) and those required for life as we know it, such as carbon, oxygen, and nitrogen. These and other elements combined in space to produce a wide variety of compounds that form the basis of life on Earth. In particular, life on Earth is based on the presence of a key unit known as an **organic molecule**, a molecule that contains carbon. Especially important are the hydrocarbons, chemical compounds made up entirely of hydrogen and carbon, which serve as the basis for our biological chemistry, or *biochemistry*. While we do not understand the details of how life on Earth began, it is clear that to make creatures like us possible, events like the ones we described must have occurred, resulting in what is called the *chemical evolution* of the universe.

What Made Earth Hospitable to Life?

About 5 billion years ago, a cloud of gas and dust in this cosmic neighborhood began to collapse under its own weight. Out of this cloud formed the Sun and its planets, together with all the smaller bodies, such as comets, that also orbit the Sun ([link]). The third planet from the Sun, as it cooled, eventually allowed the formation of large quantities of liquid water on its surface.

Comet Hyakutake.

This image was captured in 1996 by NASA photographer Bill Ingalls. Comet impacts can deliver both water and a variety of interesting chemicals, including some organic chemicals, to Earth. (credit: NASA/Bill Ingalls)

The chemical variety and moderate conditions on Earth eventually led to the formation of molecules that could make copies of themselves (reproduce), which is essential for beginning life. Over the billions of years of Earth history, life evolved and became more complex. The course of evolution was punctuated by occasional planet-wide changes caused by

collisions with some of the smaller bodies that did not make it into the Sun or one of its accompanying worlds. As we saw in the chapter on Earth as a Planet, mammals may owe their domination of Earth’s surface to just such a collision 65 million years ago, which led to the extinction of the dinosaurs (along with the majority of other living things). The details of such mass extinctions are currently the focus of a great deal of scientific interest.

Through many twisting turns, the course of evolution on Earth produced a creature with self-consciousness, able to ask questions about its own origins and place in the cosmos ([link]). Like most of Earth, this creature is composed of atoms that were forged in earlier generations of stars—in this case, assembled into both its body and brain. We might say that through the thoughts of human beings, the matter in the universe can become aware of itself.

Young Human.

Human beings have the intellect to wonder about their planet and what lies beyond it. Through them (and perhaps other intelligent life), the universe becomes aware of itself. (credit: Andrew Fraknoi)

Think about those atoms in your body for a minute. They are merely on loan to you from the lending library of atoms that make up our local corner of the universe. Atoms of many kinds circulate through your body and then leave it—with each breath you inhale and exhale and the food you eat and excrete. Even the atoms that take up more permanent residence in your tissues will not be part of you much longer than you are alive. Ultimately, you will return your atoms to the vast reservoir of Earth, where they will be incorporated into other structures and even other living things in the millennia to come.

This picture of *cosmic evolution*, of our descent from the stars, has been obtained through the efforts of scientists in many fields over many decades. Some of its details are still tentative and incomplete, but we feel reasonably confident in its broad outlines. It is remarkable how much we have been able to learn in the short time we have had the instruments to probe the physical nature of the universe.

The Copernican Principle

Our study of astronomy has taught us that we have always been wrong in the past whenever we have claimed that Earth is somehow unique. Galileo, using the newly invented technology of the telescope, showed us that Earth is not the center of the solar system, but merely one of a number of objects orbiting the Sun. Our study of the stars has demonstrated that the Sun itself is a rather undistinguished star, halfway through its long main-sequence stage like so many billions of others. There seems nothing special about our position in the Milky Way Galaxy either, and nothing surprising about our Galaxy's position in either its own group or its supercluster.

The discovery of planets around other stars confirms our idea that the formation of planets is a natural consequence of the formation of stars. We have identified thousands of exoplanets—planets orbiting around other stars, from huge ones orbiting close to their stars (informally called "hot Jupiters") down to planets smaller than Earth. A steady stream of exoplanet

discoveries is leading to the conclusion that earthlike planets occur frequently—enough that there are likely many billions of “exo-Earths” in our own Milky Way Galaxy alone. From a planetary perspective, smaller planets are not unique.

Philosophers of science sometimes call the idea that there is nothing special about our place in the universe the *Copernican principle*. Given all of the above, most scientists would be surprised if life were limited to our planet and had started nowhere else. There are billions of stars in our Galaxy old enough for life to have developed on a planet around them, and there are billions of other galaxies as well. Astronomers and biologists have long conjectured that a series of events similar to those on the early Earth probably led to living organisms on many planets around other stars, and possibly even on other planets in our solar system, such as Mars.

The real scientific issue (which we do not currently know the answer to) is whether organic biochemistry is likely or unlikely in the universe at large. Are we a fortunate and exceedingly rare outcome of chemical evolution, or is organic biochemistry a regular part of the chemical evolution of the cosmos? We do not yet know the answer to this question, but data, even an exceedingly small amount (like finding “unrelated to us” living systems on a world like Europa), will help us arrive at it.

So Where Are They?

If the Copernican principle is applied to life, then biology may be rather common among planets. Taken to its logical limit, the Copernican principle also suggests that intelligent life like us might be common. Intelligence like ours has some very special properties, including an ability to make progress through the application of technology. Organic life around other (older) stars may have started a billion years earlier than we did on Earth, so they may have had a lot more time to develop advanced technology such as sending information, probes, or even life-forms between stars.

Faced with such a prospect, physicist Enrico Fermi asked a question several decades ago that is now called the *Fermi paradox*: where are they? If life and intelligence are common and have such tremendous capacity for

growth, why is there not a network of galactic civilizations whose presence extends even into a “latecomer” planetary system like ours?

Several solutions have been suggested to the Fermi paradox. Perhaps life is common but intelligence (or at least technological civilization) is rare. Perhaps such a network will come about in the future but has not yet had the time to develop. Maybe there are invisible streams of data flowing past us all the time that we are not advanced enough or sensitive enough to detect. Maybe advanced species make it a practice not to interfere with immature, developing consciousness such as our own. Or perhaps civilizations that reach a certain level of technology then self-destruct, meaning there are no other civilizations now existing in our Galaxy. We do not yet know whether any advanced life is out there and, if it is, why we are not aware of it. Still, you might want to keep these issues in mind as you read the rest of this chapter.

Note:
Is there a network of galactic civilizations beyond our solar system? If so, why can’t we see them? Explore the possibilities in the cartoon video “The Fermi Paradox—Where Are All the Aliens?”

Key Concepts and Summary

Life on Earth is based on the presence of a key unit known as an organic molecule, a molecule that contains carbon, especially complex hydrocarbons. Our solar system formed about 5 billion years ago from a cloud of gas and dust enriched by several generations of heavier element production in stars. Life is made up of chemical combinations of these elements made by stars. The Copernican principle, which suggests that there is nothing special about our place in the universe, implies that if life could develop on Earth, it should be able to develop in other places as well. The Fermi paradox asks why, if life is common, more advanced life-forms have not contacted us.

Glossary

organic molecule
: a combination of carbon and other atoms—primarily hydrogen, oxygen, nitrogen, phosphorus, and sulfur—some of which serve as the basis for our biochemistry

Astrobiology

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the chemical building blocks required for life
- Describe the molecular systems and processes driving the origin and evolution of life
- Describe the characteristics of a habitable environment
- Describe some of the extreme conditions on Earth, and explain how certain organisms have adapted to these conditions

Scientists today take a multidisciplinary approach to studying the origin, evolution, distribution, and ultimate fate of life in the universe; this field of study is known as **astrobiology**. You may also sometimes hear this field referred to as *exobiology* or *bioastronomy.* Astrobiology brings together astronomers, planetary scientists, chemists, geologists, and biologists (among others) to work on the same problems from their various perspectives.

Among the issues that astrobiologists explore are the conditions in which life arose on Earth and the reasons for the extraordinary adaptability of life on our planet. They are also involved in identifying habitable worlds beyond Earth and in trying to understand in practical terms how to look for life on those worlds. Let’s look at some of these issues in more detail.

The Building Blocks of Life

While no unambiguous evidence for life has yet been found anywhere beyond Earth, life's chemical building blocks have been detected in a wide range of extraterrestrial environments. Meteorites (which you learned about in Cosmic Samples and the Origin of the Solar System) have been found to contain two kinds of substances whose chemical structures mark them as having an extraterrestrial origin—amino acids and sugars. **Amino acids** are **organic compounds** that are the molecular building blocks of proteins. **Proteins** are key biological molecules that provide the structure and function of the body's tissues and organs and essentially carry out the "work" of the cell. When we examine the gas and dust around comets, we also find a number of organic molecules—compounds that on Earth are associated with the chemistry of life.

Expanding beyond our solar system, one of the most interesting results of modern radio astronomy has been the discovery of organic molecules in giant clouds of gas and dust between stars. More than 100 different molecules have been identified in these reservoirs of cosmic raw material, including formaldehyde, alcohol, and others we know as important stepping stones in the development of life on Earth. Using radio telescopes and radio spectrometers, astronomers can measure the abundances of various chemicals in these clouds. We find organic molecules most readily in regions where the interstellar dust is most abundant, and it turns out these are precisely the regions where star formation (and probably planet formation) happen most easily ([link]).

Cloud of Gas and Dust.

This cloud of gas and dust in the constellation of Scorpius is the sort of region where complex molecules are found. It is also the sort of cloud where new stars form from the reservoir of gas and dust in the cloud. Radiation from a group of hot stars (off the picture to the bottom left) called the Scorpius OB Association is “eating into” the cloud, sweeping it into an elongated shape and causing the reddish glow seen at its tip. (credit: Dr. Robert Gendler)

Clearly the early Earth itself produced some of the molecular building blocks of life. Since the early 1950s, scientists have tried to duplicate in their laboratories the chemical pathways that led to life on our planet. In a series of experiments known as the *Miller-Urey experiments*, pioneered by Stanley Miller and Harold Urey at the University of Chicago, biochemists have simulated conditions on early Earth and have been able to produce some of the fundamental building blocks of life, including those that form

proteins and other large biological molecules known as nucleic acids (which we will discuss shortly).

Although these experiments produced encouraging results, there are some problems with them. The most interesting chemistry from a biological perspective takes place with hydrogen-rich or *reducing* gases, such as ammonia and methane. However, the early atmosphere of Earth was probably dominated by carbon dioxide (as Venus' and Mars' atmospheres still are today) and may not have contained an abundance of reducing gases comparable to that used in Miller-Urey type experiments. Hydrothermal vents—seafloor systems in which ocean water is superheated and circulated through crustal or mantle rocks before reemerging into the ocean—have also been suggested as potential contributors of organic compounds on the early Earth, and such sources would not require Earth to have an early reducing atmosphere.

Both earthly and extraterrestrial sources may have contributed to Earth's early supply of organic molecules, although we have more direct evidence for the latter. It is even conceivable that life itself originated elsewhere and was seeded onto our planet—although this, of course, does not solve the problem of how that life originated to begin with.

Note:
Hydrothermal vents are beginning to seem more likely as early contributors to the organic compounds found on Earth. Read about hydrothermal vents, watch videos and slideshows on these and other deep-sea wonders, and try an interactive simulation of hydrothermal circulation at the Woods Hole Oceanographic Institution website.

The Origin and Early Evolution of Life

The carbon compounds that form the chemical basis of life may be common in the universe, but it is still a giant step from these building blocks to a living cell. Even the simplest molecules of the **genes** (the basic functional

units that carry the genetic, or hereditary, material in a cell) contain millions of molecular units, each arranged in a precise sequence. Furthermore, even the most primitive life required two special capabilities: a means of extracting energy from its environment, and a means of encoding and replicating information in order to make faithful copies of itself. Biologists today can see ways that either of these capabilities might have formed in a natural environment, but we are still a long way from knowing how the two came together in the first life-forms.

We have no solid evidence for the pathway that led to the origin of life on our planet except for whatever early history may be retained in the biochemistry of modern life. Indeed, we have very little direct evidence of what Earth itself was like during its earliest history—our planet is so effective at resurfacing itself through plate tectonics (see the chapter on Earth as a Planet) that very few rocks remain from this early period. In the earlier chapter on Cratered Worlds, you learned that Earth was subjected to a heavy bombardment—a period of large impact events—some 3.8 to 4.1 billion years ago. Large impacts would have been energetic enough to heat-sterilize the surface layers of Earth, so that even if life had begun by this time, it might well have been wiped out.

When the large impacts ceased, the scene was set for a more peaceful environment on our planet. If the oceans of Earth contained accumulated organic material from any of the sources already mentioned, the ingredients were available to make living organisms. We do not understand in any detail the sequence of events that led from molecules to biology, but there is fossil evidence of microbial life in 3.5-billion-year-old rocks, and possible (debated) evidence for life as far back as 3.8 billion years.

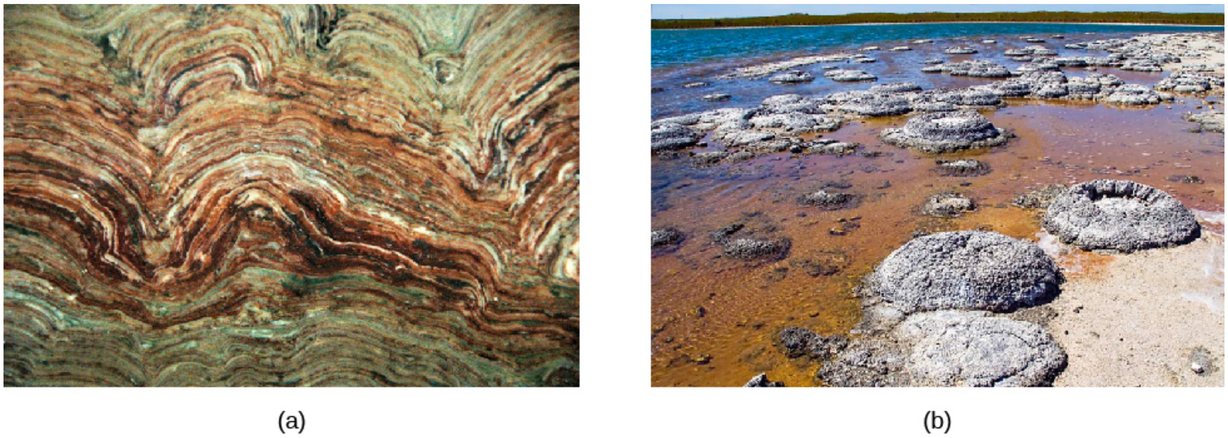
Life as we know it employs two main molecular systems: the functional molecules known as proteins, which carry out the chemical work of the cell, and information-containing molecules of **DNA (deoxyribonucleic acid)** that store information about how to create the cell and its chemical and structural components. The origin of life is sometimes considered a “chicken and egg problem” because, in modern biology, neither of these systems works without the other. It is our proteins that assemble DNA strands in the precise order required to store information, but the proteins

are created based on information stored in DNA. Which came first? Some origin of life researchers believe that prebiotic chemistry was based on molecules that could both store information and do the chemical work of the cell. It has been suggested that **RNA (ribonucleic acid)**, a molecule that aids in the flow of genetic information from DNA to proteins, might have served such a purpose. The idea of an early “RNA world” has become increasingly accepted, but a great deal remains to be understood about the origin of life.

Perhaps the most important innovation in the history of biology, apart from the origin of life itself, was the discovery of the process of **photosynthesis**, the complex sequence of chemical reactions through which some living things can use sunlight to manufacture products that store energy (such as carbohydrates), releasing oxygen as one by-product. Previously, life had to make do with sources of chemical energy available on Earth or delivered from space. But the abundant energy available in sunlight could support a larger and more productive biosphere, as well as some biochemical reactions not previously possible for life. One of these was the production of oxygen (as a waste product) from carbon dioxide, and the increase in atmospheric levels of oxygen about 2.4 billion years ago means that oxygen-producing photosynthesis must have emerged and become globally important by this time. In fact, it is likely that oxygen-producing photosynthesis emerged considerably earlier.

Some forms of chemical evidence contained in ancient rocks, such as the solid, layered rock formations known as **stromatolites**, are thought to be the fossils of oxygen-producing photosynthetic bacteria in rocks that are almost 3.5 billion years old ([link]). It is generally thought that a simpler form of photosynthesis that does not produce oxygen (and is still used by some bacteria today) probably preceded oxygen-producing photosynthesis, and there is strong fossil evidence that one or the other type of photosynthesis was functioning on Earth at least as far back as 3.4 billion years ago.

Stromatolites Preserve the Earliest Physical Representation of Life on Earth.



In their reach for sunlight, the single-celled microbes formed mats that trapped sediments in the water above them. Such trapped sediments fell and formed layers on top of the mats. The microbes then climbed atop the sediment layers and trapped more sediment. What is found in the rock record are (a) the solidified, curved sedimentary layers that are signatures of biological activity. The earliest known stromatolite is 3.47 billion years old and is found in Western Australia. (b) This more recent example is in Lake Thetis, also in Western Australia. (credit a: modification of work by James St. John; credit b: modification of work by Ruth Ellison)

The free oxygen produced by photosynthesis began accumulating in our atmosphere about 2.4 billion years ago. The interaction of sunlight with oxygen can produce ozone (which has three atoms of oxygen per molecule, as compared to the two atoms per molecule in the oxygen we breathe), which accumulated in a layer high in Earth's atmosphere. As it does on Earth today, this ozone provided protection from the Sun's damaging ultraviolet radiation. This allowed life to colonize the landmasses of our planet instead of remaining only in the ocean.

The rise in oxygen levels was deadly to some microbes because, as a highly reactive chemical, it can irreversibly damage some of the biomolecules that early life had developed in the absence of oxygen. For other microbes, it was a boon: combining oxygen with organic matter or other reduced chemicals generates a lot of energy—you can see this when a log burns, for

example—and many forms of life adopted this way of living. This new energy source made possible a great proliferation of organisms, which continued to evolve in an oxygen-rich environment.

The details of that evolution are properly the subject of biology courses, but the process of evolution by natural selection (survival of the fittest) provides a clear explanation for the development of Earth's remarkable variety of life-forms. It does not, however, directly solve the mystery of life's earliest beginnings. We hypothesize that life will arise whenever conditions are appropriate, but this hypothesis is just another form of the Copernican principle. We now have the potential to address this hypothesis with observations. If a second example of life is found in our solar system or a nearby star, it would imply that life emerges commonly enough that the universe is likely filled with biology. To make such observations, however, we must first decide where to focus our search.

Note:
Just how did life arise in the first place? And could it have happened with a different type of chemistry? Watch the 15-minute video Making Matter Come Alive in which a chemistry expert explores some answers to these questions, from a 2011 TED Talk.

Habitable Environments

Among the staggering number of objects in our solar system, Galaxy, and universe, some may have conditions suitable for life, while others do not. Understanding what conditions and features make a **habitable environment**—an environment capable of hosting life—is important both for understanding how widespread habitable environments may be in the universe and for focusing a search for life beyond Earth. Here, we discuss habitability from the perspective of the life we know. We will explore the basic requirements of life and, in the following section, consider the full range of environmental conditions on Earth where life is found. While we can't entirely rule out the possibility that other life-forms might have

biochemistry based on alternatives to carbon and liquid water, such life “as we don’t know it” is still completely speculative. In our discussion here, we are focusing on habitability for life that is chemically similar to that on Earth.

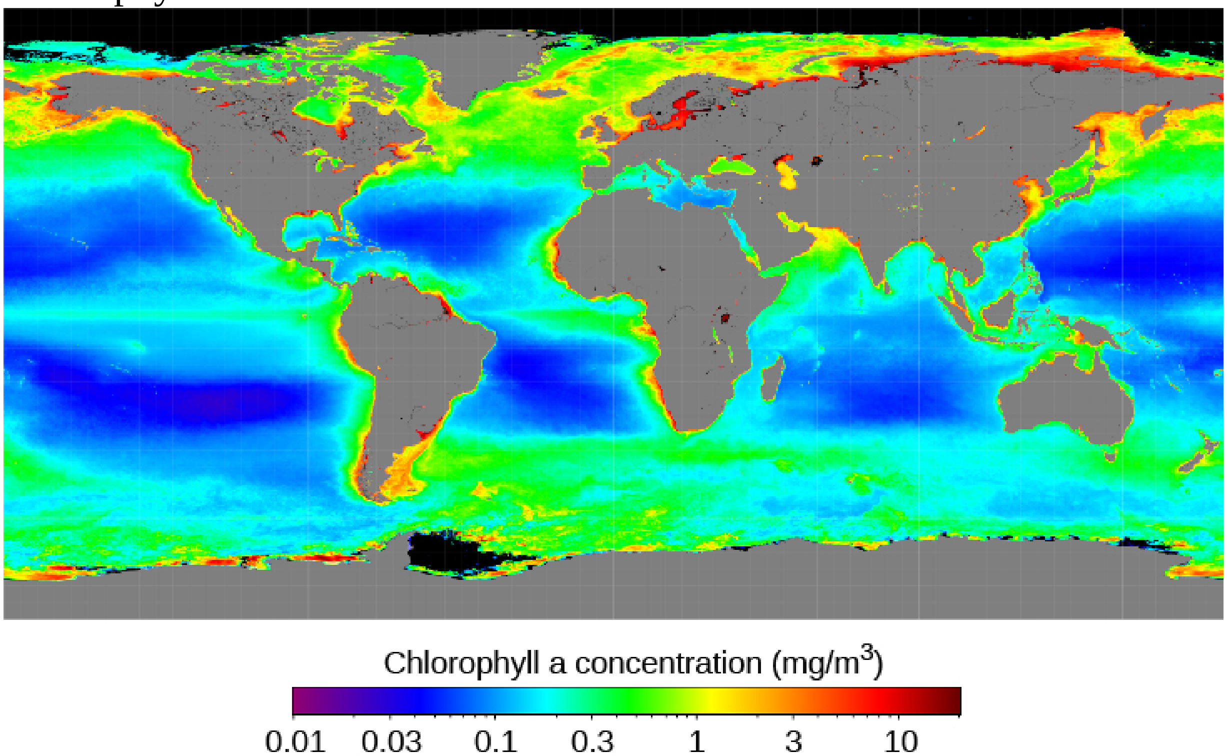
Life requires a solvent (a liquid in which chemicals can dissolve) that enables the construction of biomolecules and the interactions between them. For life as we know it, that solvent is water, which has a variety of properties that are critical to how our biochemistry works. Water is abundant in the universe, but life requires that water be in liquid form (rather than ice or gas) in order to properly fill its role in biochemistry. That is the case only within a certain range of temperatures and pressures—too high or too low in either variable, and water takes the form of a solid or a gas. Identifying environments where water is present within the appropriate range of temperature and pressure is thus an important first step in identifying habitable environments. Indeed, a “follow the water” strategy has been, and continues to be, a key driver in the exploration of planets both within and beyond our solar system.

Our biochemistry is based on molecules made of carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur. Carbon is at the core of organic chemistry. Its ability to form four bonds, both with itself and with the other elements of life, allows for the formation of a vast number of potential molecules on which to base biochemistry. The remaining elements contribute structure and chemical reactivity to our biomolecules, and form the basis of many of the interactions among them. These “biogenic elements,” sometimes referred to with the acronym CHNOPS (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur), are the raw materials from which life is assembled, and an accessible supply of them is a second requirement of habitability.

As we learned in previous chapters on nuclear fusion and the life story of the stars, carbon, nitrogen, oxygen, phosphorus, and sulfur are all formed by fusion within stars and then distributed out into their galaxy as those stars die. But how they are distributed among the planets that form within a new star system, in what form, and how chemical, physical, and geological processes on those planets cycle the elements into structures that are

accessible to biology, can have significant impacts on the distribution of life. In Earth’s oceans, for example, the abundance of phytoplankton (simple organisms that are the base of the ocean food chain) in surface waters can vary by a thousand-fold because the supply of nitrogen differs from place to place ([link]). Understanding what processes control the accessibility of elements at all scales is thus a critical part of identifying habitable environments.

Chlorophyll Abundance.



The abundance of chlorophyll (an indicator of photosynthetic bacteria and algae) varies by almost a thousand-fold across the ocean basins. That variation is almost entirely due to the availability of nitrogen—one of the major “biogenic elements” in forms that can be used by life. (credit: modification of work by NASA, Gene C. Feldman)

With these first two requirements, we have the elemental raw materials of life and a solvent in which to assemble them into the complicated molecules that drive our biochemistry. But carrying out that assembly and maintaining the complicated biochemical machinery of life takes energy. You fulfill

your own requirement for energy every time you eat food or take a breath, and you would not live for long if you failed to do either on a regular basis. Life on Earth makes use of two main types of energy: for you, these are the oxygen in the air you breathe and the organic molecules in your food. But life overall can use a much wider array of chemicals and, while all animals require oxygen, many bacteria do not. One of the earliest known life processes, which still operates in some modern microorganisms, combines hydrogen and carbon dioxide to make methane, releasing energy in the process. There are microorganisms that “breathe” metals that would be toxic to us, and even some that breathe in sulfur and breathe out sulfuric acid. Plants and photosynthetic microorganisms have also evolved mechanisms to use the energy in light directly.

Water in the liquid phase, the biogenic elements, and energy are the fundamental requirements for habitability. But are there additional environmental constraints? We consider this in the next section.
Grand Prismatic Spring in Yellowstone National Park.

This hot spring, where water emerges from the bluish center at temperatures near the local boiling point

(about 92 ºC), supports a thriving array of microbial life. The green, yellow, and orange colors around the edges come from thick “mats” of photosynthetic bacteria. In fact, their coloration in part demonstrates their use of light energy—some wavelengths of incoming sunlight are selectively captured for energy; the rest are reflected back. Since it lacks the captured wavelengths, this light is now different in color than the sunlight that illuminates it. The blue part of the spring has temperatures too high to allow photosynthetic life (hence the lack of color except that supplied by water itself), but life is still present. Here, at nearly boiling temperatures, bacteria use the chemical energy supplied by the combination of hydrogen and other chemicals with oxygen. (credit: modification of work by Domenico Salvagnin)

Life in Extreme Conditions

At a chemical level, life consists of many types of molecules that interact with one another to carry out the processes of life. In addition to water, elemental raw materials, and energy, life also needs an environment in which those complicated molecules are stable (don’t break down before they can do their jobs) and their interactions are possible. Your own biochemistry works properly only within a very narrow range of about 10 °C in body temperature and two-tenths of a unit in blood pH (pH is a numerical measure of acidity, or the amount of free hydrogen ions). Beyond those limits, you are in serious danger.

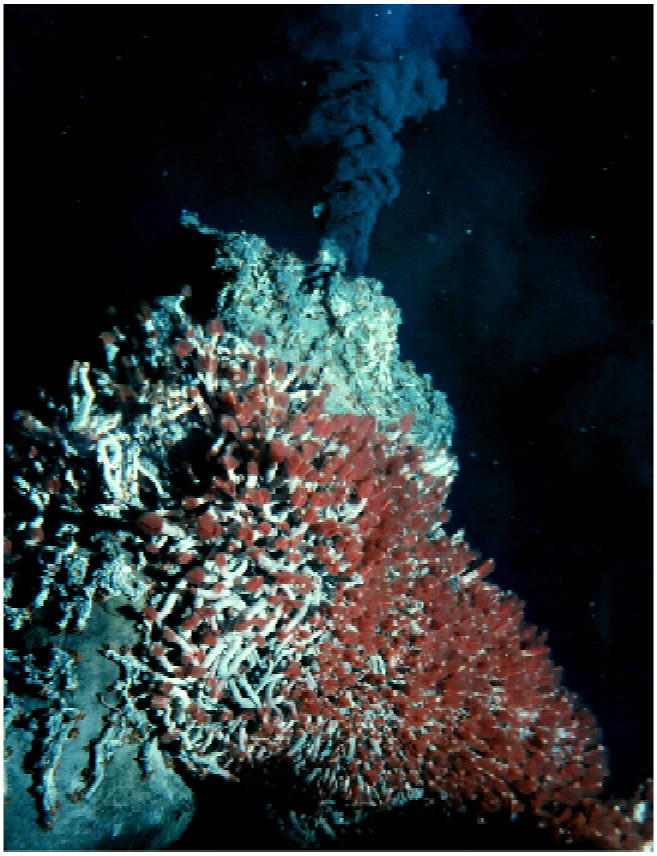
Life overall must also have limits to the conditions in which it can properly work but, as we will see, they are much broader than human limits. The resources that fuel life are distributed across a very wide range of conditions. For example, there is abundant chemical energy to be had in hot springs that are essentially boiling acid (see [link]). This provides ample incentive for evolution to fill as much of that range with life as is

biochemically possible. An organism (usually a microbe) that tolerates or even thrives under conditions that most of the life around us would consider hostile, such as very high or low temperature or acidity, is known as an **extremophile** (where the suffix *-phile* means “lover of”). Let’s have a look at some of the conditions that can challenge life and the organisms that have managed to carve out a niche at the far reaches of possibility.

Both high and low temperatures can cause a problem for life. As a large organism, you are able to maintain an almost constant body temperature whether it is colder or warmer in the environment around you. But this is not possible at the tiny size of microorganisms; whatever the temperature in the outside world is also the temperature of the microbe, and its biochemistry must be able to function at that temperature. High temperatures are the enemy of complexity—increasing thermal energy tends to break apart big molecules into smaller and smaller bits, and life needs to stabilize the molecules with stronger bonds and special proteins. But this approach has its limits.

Nevertheless, as noted earlier, high-temperature environments like hot springs and hydrothermal vents often offer abundant sources of chemical energy and therefore drive the evolution of organisms that can tolerate high temperatures (see [link]); such an organism is called a **thermophile**. Currently, the high temperature record holder is a methane-producing microorganism that can grow at 122 °C, where the pressure also is so high that water still does not boil. That’s amazing when you think about it. We cook our food—meaning, we alter the chemistry and structure of its biomolecules—by boiling it at a temperature of 100 °C. In fact, food begins to cook at much lower temperatures than this. And yet, there are organisms whose biochemistry remains intact and operates just fine at temperatures 20 degrees higher.

Hydrothermal Vent on the Sea Floor.



What appears to be black smoke is actually superheated water filled with minerals of metal sulfide. Hydrothermal vent fluid can represent a rich source of chemical energy, and therefore a driver for the evolution of microorganisms that can tolerate high temperatures. Bacteria feeding on this chemical energy form the base of a food chain that can support thriving communities of animals—in this case, a dense patch of red and white tubeworms growing around the base of the vent. (credit: modification of work by the University of Washington; NOAA/OAR/OER)

Cold can also be a problem, in part because it slows down metabolism to very low levels, but also because it can cause physical changes in biomolecules. Cell membranes—the molecular envelopes that surround cells and allow their exchange of chemicals with the world outside—are basically made of fatlike molecules. And just as fat congeals when it cools, membranes crystallize, changing how they function in the exchange of materials in and out of the cell. Some cold-adapted cells (called *psychrophiles*) have changed the chemical composition of their membranes in order to cope with this problem; but again, there are limits. Thus far, the coldest temperature at which any microbe has been shown to reproduce is about –25 ºC.

Conditions that are very acidic or alkaline can also be problematic for life because many of our important molecules, like proteins and DNA, are broken down under such conditions. For example, household drain cleaner, which does its job by breaking down the chemical structure of things like hair clogs, is a very alkaline solution. The most acid-tolerant organisms (*acidophiles*) are capable of living at pH values near zero—about ten million times more acidic than your blood ([link]). At the other extreme, some *alkaliphiles* can grow at pH levels of about 13, which is comparable to the pH of household bleach and almost a million times more alkaline than your blood.

Spain’s Rio Tinto.

With a pH close to 2, Rio Tinto is literally a river of acid. Acid-loving microorganisms (acidophiles) not only thrive in these waters, their metabolic activities help generate the acid in the first place. The rusty red color that gives the river its name comes from high levels of iron dissolved in the waters.

High levels of salts in the environment can also cause a problem for life because the salt blocks some cellular functions. Humans recognized this centuries ago and began to salt-cure food to keep it from spoiling—meaning, to keep it from being colonized by microorganisms. Yet some microbes have evolved to grow in water that is saturated in sodium chloride (table salt)—about ten times as salty as seawater ([link]).

Salt Ponds.

The waters of an evaporative salt works near San Francisco are colored pink by thriving communities of photosynthetic organisms. These waters are about ten times as salty as seawater—enough for sodium chloride to begin to crystallize out—yet some organisms can survive and thrive in these conditions. (credit: modification of work by NASA)

Very high pressures can literally squeeze life's biomolecules, causing them to adopt more compact forms that do not work very well. But we still find life—not just microbial, but even animal life—at the bottoms of our ocean trenches, where pressures are more than 1000 times atmospheric pressure. Many other adaptions to environmental "extremes" are also known. There is even an organism, *Deinococcus radiodurans*, that can tolerate ionizing radiation (such as that released by radioactive elements) a thousand times more intense than you would be able to withstand. It is also very good at

surviving extreme desiccation (drying out) and a variety of metals that would be toxic to humans.

From many such examples, we can conclude that life is capable of tolerating a wide range of environmental extremes—so much so that we have to work hard to identify places where life can't exist. A few such places are known—for example, the waters of hydrothermal vents at over 300 °C appear too hot to support any life—and finding these places helps define the possibility for life elsewhere. The study of extremophiles over the last few decades has expanded our sense of the range of conditions life can survive and, in doing so, has made many scientists more optimistic about the possibility that life might exist beyond Earth.

Key Concepts and Summary

The study of life in the universe, including its origin on Earth, is called astrobiology. Life as we know it requires water, certain elemental raw materials (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur), energy, and an environment in which the complex chemistry of life is stable. Carbon-based (or organic) molecules are abundant in space and may also have been produced by processes on Earth. Life appears to have spread around our planet within 400 million years after the end of heavy bombardment, if not sooner. The actual origin of life—the processes leading from chemistry to biology—is not completely understood. Once life took hold, it evolved to use many energy sources, including first a range of different chemistries and later light, and diversified across a range of environmental conditions that humans consider "extreme." This proliferation of life into so many environmental niches, so relatively soon after our planet became habitable, has served to make many scientists optimistic about the chances that life could exist elsewhere.

Glossary

amino acids
: organic compounds that are the molecular building blocks of proteins

astrobiology

the multidisciplinary study of life in the universe: its origin, evolution, distribution, and fate; similar terms are *exobiology* and *bioastronomy*

DNA (deoxyribonucleic acid)
: a molecule that stores information about how to replicate a cell and its chemical and structural components

extremophile
: an organism (usually a microbe) that tolerates or even thrives under conditions that most of the life around us would consider hostile, such as very high or low temperature or acidity

gene
: the basic functional unit that carries the genetic (hereditary) material contained in a cell

habitable environment
: an environment capable of hosting life

organic compound
: a compound containing carbon, especially a complex carbon compound; not necessarily produced by life

photosynthesis
: a complex sequence of chemical reactions through which some living things can use sunlight to manufacture products that store energy (such as carbohydrates), releasing oxygen as one by-product

protein
: a key biological molecule that provides the structure and function of the body’s tissues and organs, and essentially carries out the chemical work of the cell

RNA (ribonucleic acid)
: a molecule that aids in the flow of genetic information from DNA to proteins

stromatolites

solid, layered rock formations that are thought to be the fossils of oxygen-producing photosynthetic bacteria in rocks that are 3.5 billion years old

thermophile
an organism that can tolerate high temperatures

Searching for Life beyond Earth

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Outline what we have learned from exploration of the environment on Mars
- Identify where in the solar system life is most likely sustainable and why
- Describe some key missions and their findings in our search for life beyond our solar system
- Explain the use of biomarkers in the search for evidence of life beyond our solar system

Astronomers and planetary scientists continue to search for life in the solar system and the universe at large. In this section, we discuss two kinds of searches. First is the direct exploration of planets within our own solar system, especially Mars and some of the icy moons of the outer solar system. Second is the even more difficult task of searching for evidence of life—a **biomarker**—on planets circling other stars. In the next section, we will examine SETI, the *search for extraterrestrial intelligence*. As you will see, the approaches taken in these three cases are very different, even though the goal of each is the same: to determine if life on Earth is unique in the universe.

Life on Mars

The possibility that Mars hosts, or has hosted, life has a rich history dating back to the “canals” that some people claimed to see on the martian surface toward the end of the nineteenth century and the beginning of the twentieth. With the dawn of the space age came the possibility to address this question up close through a progression of missions to Mars that began with the first successful flyby of a robotic spacecraft in 1964 and have led to the deployment of NASA’s *Curiosity* rover, which landed on Mars’ surface in 2012.

The earliest missions to Mars provided some hints that liquid water—one of life’s primary requirements—may once have flowed on the surface, and later missions have strengthened this conclusion. The NASA Viking landers, whose purpose was to search directly for evidence of life on Mars, arrived on Mars in 1976. Viking’s onboard instruments found no organic molecules (the stuff of which life is made), and no evidence of biological activity in the martian soils it analyzed.

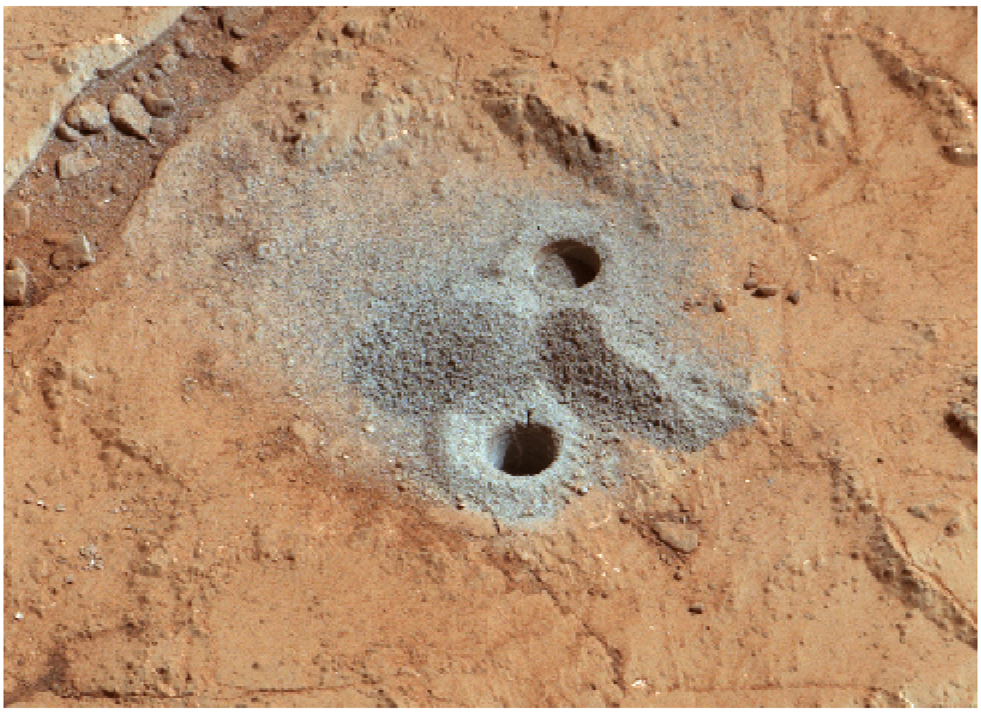
This result is not particularly surprising because, despite the evidence of flowing liquid water in the past, liquid water on the surface of Mars is generally not stable today. Over much of Mars, temperatures and pressures at the surface are so low that pure water would either freeze or boil away (under very low pressures, water will boil at a much lower temperature than usual). To make matters worse, unlike Earth, Mars does not have a magnetic field and ozone layer to protect the surface from harmful solar ultraviolet radiation and energetic particles. However, Viking’s analyses of the soil said nothing about whether life may have existed in Mars’ distant past, when liquid water was more abundant. We do know that water in the form of ice exists in abundance on Mars, not so deep beneath its surface. Water vapor is also a constituent of the atmosphere of Mars.

Since the visit of Viking, our understanding of Mars has deepened spectacularly. Orbiting spacecraft have provided ever-more detailed images of the surface and detected the presence of minerals that could have formed only in the presence of liquid water. Two bold surface missions, the Mars Exploration Rovers *Spirit* and *Opportunity* (2004), followed by the much larger *Curiosity* Rover (2012), confirmed these remote-sensing data. All three rovers found abundant evidence for a past history of liquid water,

revealed not only from the mineralogy of rocks they analyzed, but also from the unique layering of rock formations.

Curiosity has gone a step beyond evidence for water and confirmed the existence of habitable environments on ancient Mars. “Habitable” means not only that liquid water was present, but that life’s requirements for energy and elemental raw materials could also have been met. The strongest evidence of an ancient habitable environment came from analyzing a very fine-grained rock called a mudstone—a rock type that is widespread on Earth but was unknown on Mars until *Curiosity* found it (see [link]). The mudstone can tell us a great deal about the wet environments in which they formed.

Mudstone.



Shown are the first holes drilled by NASA’s *Curiosity* Mars rover into a mudstone, with “fresh” drill-pilings around the holes. Notice the difference in color between the red ancient martian surface and the gray newly exposed rock powder that came from the drill holes. Each drill hole is about 0.6 inch (1.6 cm) in diameter. (credit: modification of work by NASA/JPL-Caltech/MSSS)

Five decades of robotic exploration have allowed us to develop a picture of how Mars evolved through time. Early Mars had epochs of warmer and wetter conditions that would have been conducive to life at the surface. However, Mars eventually lost much of its early atmosphere and the surface water began to dry up. As that happened, the ever-shrinking reservoirs of liquid water on the martian surface became saltier and more acidic, until the surface finally had no significant liquid water and was bathed in harsh solar radiation. The surface thus became uninhabitable, but this might not be the case for the planet overall.

Reservoirs of ice and liquid water could still exist underground, where pressure and temperature conditions make it stable. There is recent evidence to suggest that liquid water (probably very salty water) can occasionally (and briefly) flow on the surface even today. Thus, Mars might even have habitable conditions in the present day, but of a much different sort than we normally think of on Earth.

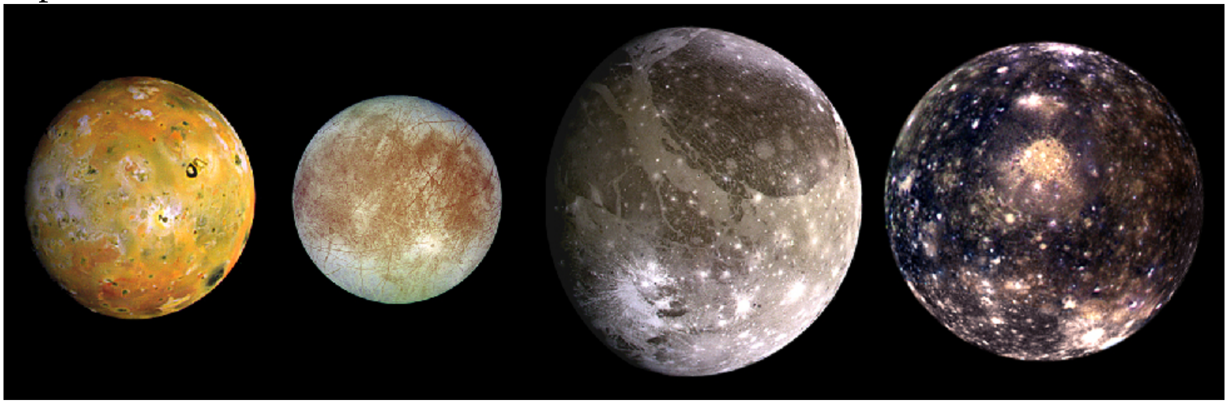
Our study of Mars reveals a planet with a fascinating history—one that saw its ability to host surface life dwindle billions of years ago, but perhaps allowing life to adapt and survive in favorable environmental niches. Even if life did not survive, we expect that we might find evidence of life if it ever took hold on Mars. If it is there, it is hidden in the crust, and we are still learning how best to decipher that evidence.

Life in the Outer Solar System

The massive gas and ice giant planets of the outer solar system—Jupiter, Saturn, Uranus, and Neptune—are almost certainly not habitable for life as we know it, but some of their moons might be (see [link]). Although these worlds in the outer solar system contain abundant water, they receive so little warming sunlight in their distant orbits that it was long believed they would be "geologically dead" balls of hard-frozen ice and rock. But, as we saw in the chapter on Rings, Moons, and Pluto, missions to the outer solar system have found something much more interesting.

Jupiter’s moon Europa revealed itself to the Voyager and Galileo missions as an active world whose icy surface apparently conceals an ocean with a depth of tens to perhaps a hundred kilometers. As the moon orbits Jupiter, the planet’s massive gravity creates tides on Europa—just as our own Moon’s gravity creates our ocean tides—and the friction of all that pushing and pulling generates enough heat to keep the water in liquid form ([link]). Similar tides act upon other moons if they orbit close to the planet. Scientists now think that six or more of the outer solar system’s icy moons may harbor liquid water oceans for the same reason. Among these, Europa and Enceladus, a moon of Saturn, have thus far been of greatest interest to astrobiologists.

Jupiter’s Moons.

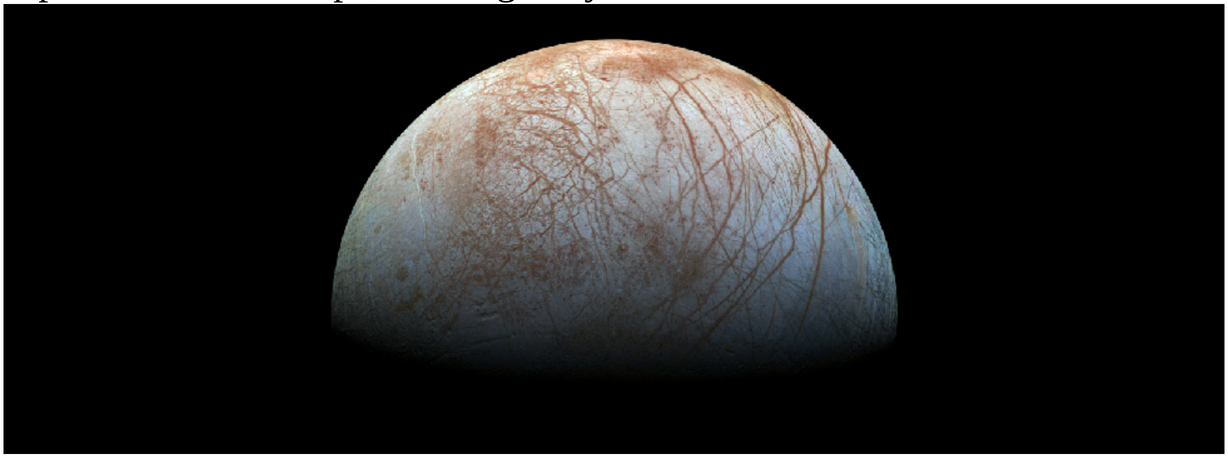


The Galilean moons of Jupiter are shown to relative scale and arranged in order of their orbital distance from Jupiter. At far left, Io orbits closest to Jupiter and so experiences the strongest tidal heating by Jupiter’s massive gravity. This effect is so strong that Io is thought to be the most volcanically active body in our solar system. At far right, Callisto shows a surface scarred by billions of years’ worth of craters—an indication that the moon’s surface is old and that Callisto may be far less active than its sibling moons. Between these hot and cold extremes, Europa, second from left, orbits at a distance where Jupiter’s tidal heating may be “just right” to sustain a liquid water ocean beneath its icy crust. (credit: modification of work by NASA/JPL/DLR)

Europa has probably had an ocean for most or all of its history, but habitability requires more than just liquid water. Life also requires energy, and because sunlight does not penetrate below the kilometers-thick ice crust of Europa, this would have to be chemical energy. One of Europa's key attributes from an astrobiology perspective is that its ocean is most likely in direct contact with an underlying rocky mantle, and the interaction of water and rocks—especially at high temperatures, as within Earth's hydrothermal vent systems—yields a *reducing chemistry* (where molecules tend to give up electrons readily) that is like one half of a chemical battery. To complete the battery and provide energy that could be used by life requires that an *oxidizing chemistry* (where molecules tend to accept electrons readily) also be available. On Earth, when chemically reducing vent fluids meet oxygen-containing seawater, the energy that becomes available often supports thriving communities of microorganisms and animals on the sea floor, far from the light of the Sun.

The Galileo mission found that Europa's icy surface does contain an abundance of oxidizing chemicals. This means that availability of energy to support life depends very much on whether the chemistry of the surface and the ocean can mix, despite the kilometers of ice in between. That Europa's ice crust appears geologically "young" (only tens of millions of years old, on average) and that it is active makes it tantalizing to think that such mixing might indeed occur. Understanding whether and how much exchange occurs between the surface and ocean of Europa will be a key science objective of future missions to Europa, and a major step forward in understanding whether this moon could be a cradle of life.

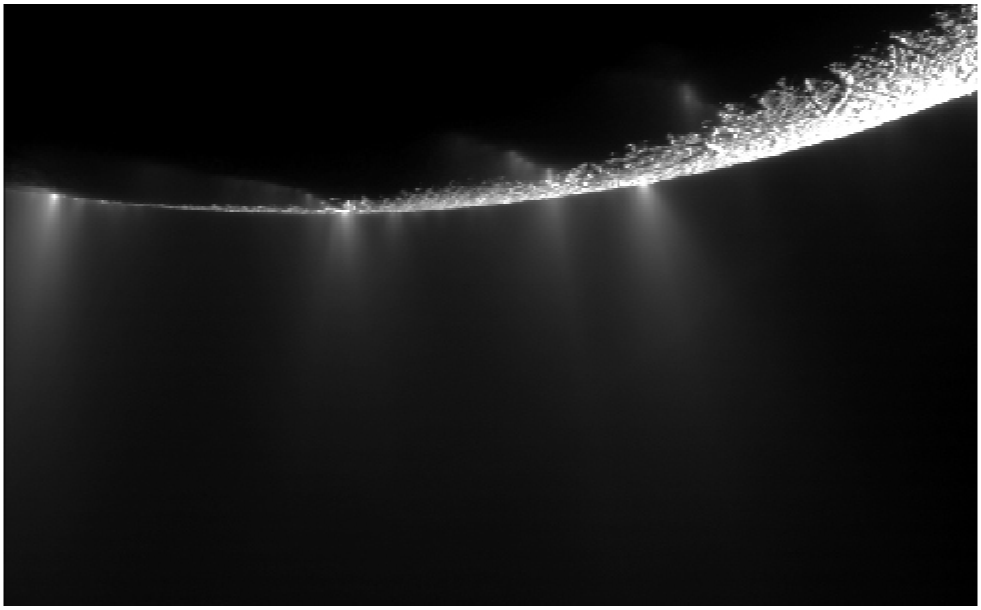
Jupiter's Moon Europa, as Imaged by NASA's Galileo Mission.



The relative scarcity of craters on Europa suggests a surface that is "geologically young," and the network of colored ridges and cracks suggests constant activity and motion. Galileo's instruments also strongly suggested the presence of a massive ocean of salty liquid water beneath the icy crust. (credit: modification of work by NASA/JPL-Caltech/SETI Institute)

In 2005, the Cassini mission performed a close flyby of a small (500-kilometer diameter) moon of Saturn, Enceladus ([link]), and made a remarkable discovery. Plumes of gas and icy material were venting from the moon's south polar region at a collective rate of about 250 kilograms of material per second. Several observations, including the discovery of salts associated with the icy material, suggest that their source is a liquid water ocean beneath tens of kilometers of ice. Although it remains to be shown definitively whether the ocean is local or global, transient or long-lived, it does appear to be in contact, and to have reacted, with a rocky interior. As on Europa, this is probably a necessary—though not sufficient—condition for habitability. What makes Enceladus so enticing to planetary scientists, though, are those plumes of material that seem to come directly from its ocean: samples of the interior are there for the taking by any spacecraft sent flying through. For a future mission, such samples could yield evidence not only of whether Enceladus is habitable but, indeed, of whether it is home to life.

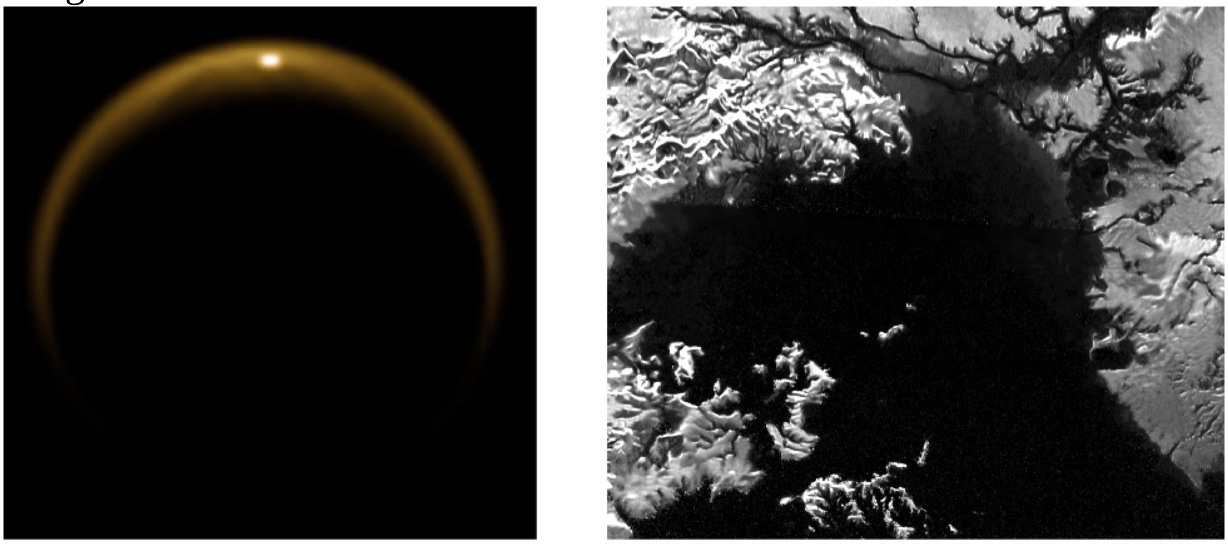
Image of Saturn's Moon Enceladus from NASA's Cassini Mission.



The south polar region was found to have multiple plumes of ice and gas that, combined, are venting about 250 kilograms of material per second into space. Such features suggest that Enceladus, like Europa, has a sub-ice ocean. (credit: NASA/JPL/ SSI)

Saturn's big moon Titan is very different from both Enceladus and Europa (see [link]). Although it may host a liquid water layer deep within its interior, it is the surface of Titan and its unusual chemistry that makes this moon such an interesting place. Titan's thick atmosphere—the only one among moons in the solar system—is composed mostly of nitrogen but also of about 5% methane. In the upper atmosphere, the Sun's ultraviolet light breaks apart and recombines these molecules into more complex organic compounds that are collectively known as *tholins*. The tholins shroud Titan in an orange haze, and imagery from Cassini and from the Huygens probe that descended to Titan's surface show that heavier particles appear to accumulate on the surface, even forming "dunes" that are cut and sculpted by flows of liquid hydrocarbons (such as liquid methane). Some scientists see this organic chemical factory as a natural laboratory that may yield some clues about the solar system's early chemistry—perhaps even chemistry that could support the origin of life.

Image of Saturn’s Moon Titan from NASA’s Cassini Mission.



(a) (b)

(a) The hazy orange glow comes from Titan’s thick atmosphere (the only one known among the moons of the solar system). That atmosphere is mostly nitrogen but also contains methane and potentially a variety of complex organic compounds. The bright spot near the top of the image is sunlight reflected from a very flat surface—almost certainly a liquid. We see this effect, called “glint,” when sunlight reflects off the surface of a lake or ocean. (b) Cassini radar imagery shows what look very much like landforms and lakes on the surface of Titan. But the surface lakes and oceans of Titan are not water; they are probably made of liquid hydrocarbons like methane and ethane. (credit a: modification of work by NASA/JPL/University of Arizona/DLR; credit b: modification of work by NASA/JPL-Caltech/ASI)

Note:
In January 2005, the Huygens probe descended to the surface of Titan and relayed data, including imagery of the landing site, for about 90 minutes. You can watch a video about the descent of Huygens to Titan’s surface.

Habitable Planets Orbiting Other Stars

One of the most exciting developments in astronomy during the last two decades is the ability to detect exoplanets—planets orbiting other stars. As we saw in the chapter on the formation of stars and planets, since the discovery of the first exoplanet in 1995, there have been thousands of confirmed detections, and many more candidates that are not yet confirmed. These include several dozen possibly habitable exoplanets. Such numbers finally allow us to make some predictions about exoplanets and their life-hosting potential. The majority of stars with mass similar to the Sun appear to host at least one planet, with multi-planet systems like our own not unusual. How many of these planets might be habitable, and how could we search for life there?

Note:
The NASA Exoplanet Archive is an up-to-date searchable online source of data and tools on everything to do with exoplanets. Explore stellar and exoplanet parameters and characteristics, find the latest news on exoplanet discoveries, plot your own data interactively, and link to other related resources.

In evaluating the prospect for life in distant planetary systems, astrobiologists have developed the idea of a **habitable zone**—a region around a star where suitable conditions might exist for life. This concept focuses on life’s requirement for liquid water, and the habitable zone is generally thought of as the range of distances from the central star in which water could be present in liquid form at a planet’s surface. In our own solar system, for example, Venus has surface temperatures far above the boiling point of water and Mars has surface temperatures that are almost always below the freezing point of water. Earth, which orbits between the two, has a surface temperature that is “just right” to keep much of our surface water in liquid form.

Whether surface temperatures are suitable for maintaining liquid water depends on a planet's "radiation budget" —how much starlight energy it absorbs and retains—and whether or how processes like winds and ocean circulation distribute that energy around the planet. How much stellar energy a planet receives, in turn, depends on how much and what sort of light the star emits and how far the planet is from that star,[footnote] how much it reflects back to space, and how effectively the planet's atmosphere can retain heat through the greenhouse effect (see Earth as a Planet). All of these can vary substantially, and all matter a lot. For example, Venus receives about twice as much starlight per square meter as Earth but, because of its dense cloud cover, also reflects about twice as much of that light back to space as Earth does. Mars receives only about half as much starlight as Earth, but also reflects only about half as much. Thus, despite their differing orbital distances, the three planets actually absorb comparable amounts of sunlight energy. Why, then, are they so dramatically different?

The amount of starlight received per unit area of a planet's surface (per square meter, for example) decreases with the square of the distance from the star. Thus, when the orbital distance doubles, the illumination decreases by 4 times (2^2), and when the orbital distance increases tenfold, the illumination decreases by 100 times (10^2). Venus and Mars orbit the sun at about 72% and 152% of Earth's orbital distance, respectively, so Venus receives about $1/(0.72)^2 = 1.92$ (about twice) and Mars about $1/(1.52)^2 = 0.43$ (about half) as much light per square meter of planet surface as Earth does.

As we learned in several chapters about the planets, some of the gases that make up planetary atmospheres are very effective at trapping infrared light —the very range of wavelengths at which planets radiate thermal energy back out to space—and this can raise the planet's surface temperature quite a bit more than would otherwise be the case. This is the same "greenhouse effect" that is of such concern for global warming on our planet. Earth's natural greenhouse effect, which comes mostly from water vapor and carbon dioxide in the atmosphere, raises our average surface temperature by about 33 °C over the value it would have if there were no greenhouse gases in the atmosphere. Mars has a very thin atmosphere and thus very little greenhouse warming (about 2 °C worth), while Venus has a massive carbon

dioxide atmosphere that creates very strong greenhouse warming (about 510 °C worth). These worlds are much colder and much hotter, respectively, than Earth would be if moved into their orbits. Thus, we must consider the nature of any atmosphere as well as the distance from the star in evaluating the range of habitability.

Of course, as we have learned, stars also vary widely in the intensity and spectrum (the wavelengths of light) they emit. Some are much brighter and hotter (bluer), while others are significantly dimmer and cooler (redder), and the distance of the habitable zone varies accordingly. For example, the habitable zone around M-dwarf stars is 3 to 30 times closer in than for G-type (Sun-like) stars. There is a lot of interest in whether such systems could be habitable because—although they have some potential downsides for supporting life—M-dwarf stars are by far the most numerous and long-lived in our Galaxy.

The luminosity of stars like the Sun also increases over their main-sequence lifetime, and this means that the habitable zone migrates outward as a star system ages. Calculations indicate that the power output of the Sun, for example, has increased by at least 30% over the past 4 billion years. Thus, Venus was once within the habitable zone, while Earth received a level of solar energy insufficient to keep the modern Earth (with its present atmosphere) from freezing over. In spite of this, there is plenty of geological evidence that liquid water was present on Earth’s surface billions of years ago. The phenomenon of increasing stellar output and an outwardly migrating habitable zone has led to another concept: the *continuously* habitable zone is defined by the range of orbits that would remain within the habitable zone during the entire lifetime of the star system. As you might imagine, the continuously habitable zone is quite a bit narrower than the habitable zone is at any one time in a star’s history. The nearest star to the Sun, Proxima Centauri, is an M star that has a planet with a mass of at least 1.3 Earth masses, taking about 11 days to orbit. At the distance for such a quick orbit (0.05 AU), the planet may be in the habitable zone of its star, although whether conditions on such a planet near such a star are hospitable for life is a matter of great scientific debate.

Even when planets orbit within the habitable zone of their star, it is no guarantee that they are habitable. For example, Venus today has virtually no water, so even if it were suddenly moved to a "just right" orbit within the habitable zone, a critical requirement for life would still be lacking.

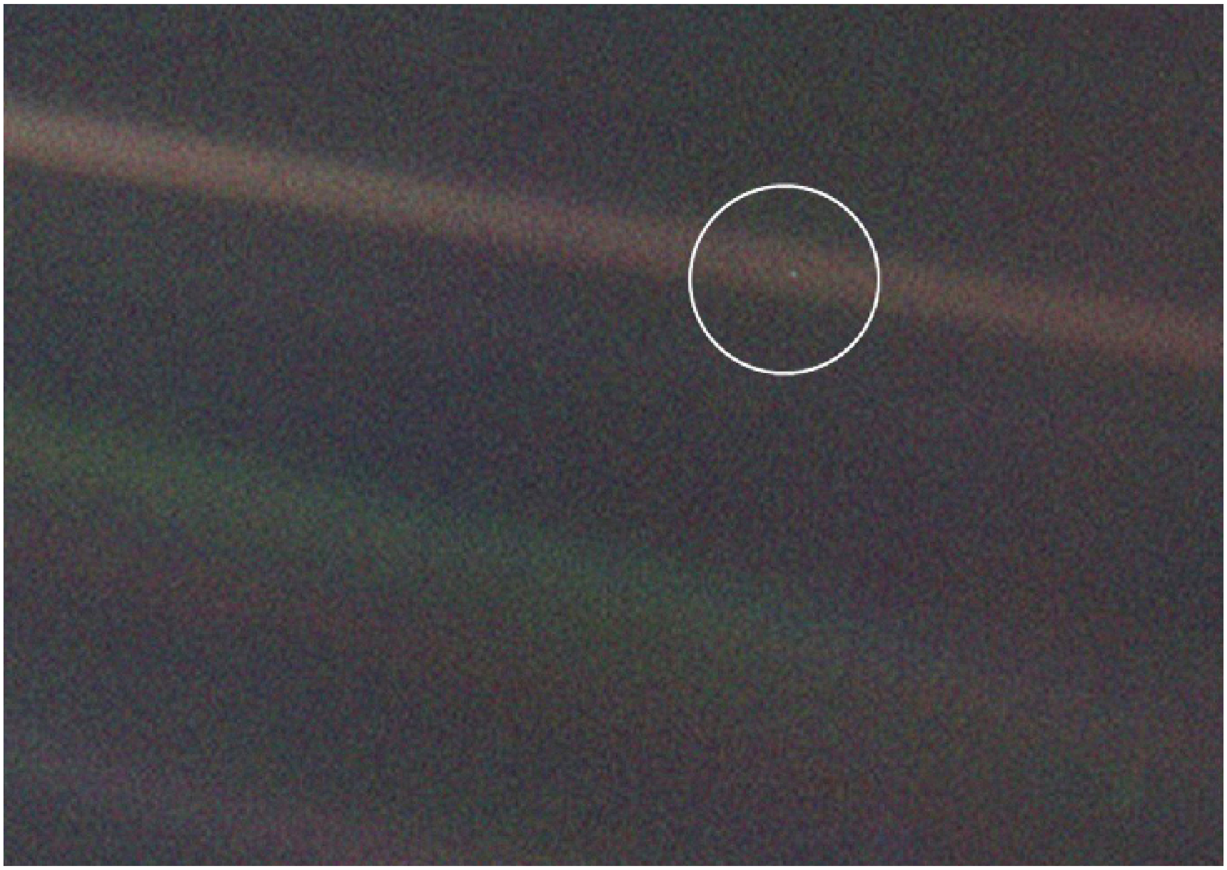
Scientists are working to understand all the factors that define the habitable zone and the habitability of planets orbiting within that zone because this will be our primary guide in targeting exoplanets on which to seek evidence of life. As technology for detecting exoplanets has advanced, so too has our potential to find Earth-size worlds within the habitable zones of their parent stars. Of the confirmed or candidate exoplanets known at the time of writing, nearly 300 are considered to be orbiting within the habitable zone and more than 10% of those are roughly Earth-size.

> **Note:**
> Explore the habitable universe at the online Planetary Habitability Laboratory created by the University of Puerto Rico at Arecibo. See the potentially habitable exoplanets and other interesting places in the universe, watch video clips, and link to numerous related resources on astrobiology.

Biomarkers

Our observations suggest increasingly that Earth-size planets orbiting within the habitable zone may be common in the Galaxy—current estimates suggest that more than 40% of stars have at least one. But are any of them inhabited? With no ability to send probes there to sample, we will have to derive the answer from the light and other radiation that come to us from these faraway systems ([link]). What types of observations might constitute good evidence for life?

Earth, as Seen by NASA's Voyager 1.



In this image, taken from 4 billion miles away, Earth appears as a "pale blue dot" representing less than a pixel's worth of light. Would this light reveal Earth as a habitable and inhabited world? Our search for life on exoplanets will depend on an ability to extract information about life from the faint light of faraway worlds. (credit: modification of work by NASA/JPL-Caltech)

To be sure, we need to look for robust biospheres (atmospheres, surfaces, and/or oceans) capable of creating planet-scale change. Earth hosts such a biosphere: the composition of our atmosphere and the spectrum of light reflected from our planet differ considerably from what would be expected in the absence of life. Presently, Earth is the only body in our solar system for which this is true, despite the possibility that habitable conditions might prevail in the subsurface of Mars or inside the icy moons of the outer solar system. Even if life exists on these worlds, it is very unlikely that it could

yield planet-scale changes that are both telescopically observable and clearly biological in origin.

What makes Earth “special” among the potentially habitable worlds in our solar system is that it has a photosynthetic biosphere. This requires the presence of liquid water at the planet’s surface, where organisms have direct access to sunlight. The habitable zone concept focuses on this requirement for surface liquid water—even though we know that subsurface habitable conditions could prevail at more distant orbits—exactly because these worlds would have biospheres detectable at a distance.

Indeed, plants and photosynthetic microorganisms are so abundant at Earth’s surface that they affect the color of the light that our planet reflects out into space—we appear greener in visible wavelengths and reflect more near-infrared light than we otherwise would. Moreover, photosynthesis has changed Earth’s atmosphere at a large scale—more than 20% of our atmosphere comes from the photosynthetic waste product, oxygen. Such high levels would be very difficult to explain in the absence of life. Other gases, such as nitrous oxide and methane, when found simultaneously with oxygen, have also been suggested as possible indicators of life. When sufficiently abundant in an atmosphere, such gases could be detected by their effect on the spectrum of light that a planet emits or reflects. (As we saw in the chapter on exoplanets, astronomers today are beginning to have the capability of detecting the spectrum of the atmospheres of some planets orbiting other stars.)

Astronomers have thus concluded that, at least initially, a search for life outside our solar system should focus on exoplanets that are as much like Earth as possible—roughly Earth-size planets orbiting in the habitable zone—and look for the presence of gases in the atmosphere or colors in the visible spectrum that are hard to explain except by the presence of biology. Simple, right? In reality, the search for exoplanet life poses many challenges.

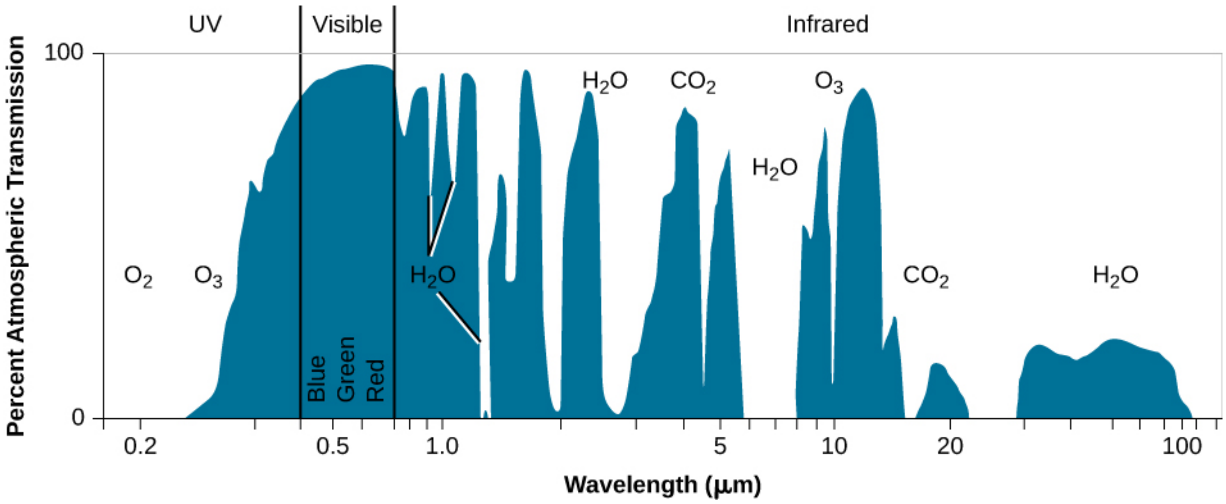
As you might imagine, this task is more challenging for planetary systems that are farther away and, in practical terms, this will limit our search to the habitable worlds closest to our own. Should we become limited to a very

small number of nearby targets, it will also become important to consider the habitability of planets orbiting the M-dwarfs we discussed above.

If we manage to separate out a clean signal from the planet and find some features in the light spectrum that might be indicative of life, we will need to work hard to think of any nonbiological process that might account for them. “Life is the hypothesis of last resort,” noted astronomer Carl Sagan—meaning that we must exhaust all other explanations for what we see before claiming to have found evidence of extraterrestrial biology. This requires some understanding of what processes might operate on worlds that we will know relatively little about; what we find on Earth can serve as a guide but also has potential to lead us astray ([link]).

Recall, for example, that it would be extremely difficult to account for the abundance of oxygen in Earth’s atmosphere except by the presence of biology. But it has been hypothesized that oxygen could build up to substantial levels on planets orbiting M-dwarf stars through the action of ultraviolet radiation on the atmosphere—with no need for biology. It will be critical to understand where such “false positives” might exist in carrying out our search.

We need to understand that we might not be able to detect biospheres even if they exist. Life has flourished on Earth for perhaps 3.5 billion years, but the atmospheric “biosignatures” that, today, would supply good evidence for life to distant astronomers have not been present for all of that time. Oxygen, for example, accumulated to detectable levels in our atmosphere only a little over 2 billion years ago. Could life on Earth have been detected before that time? Scientists are working actively to understand what additional features might have provided evidence of life on Earth during that early history, and thereby help our chances of finding life beyond.
Spectrum of Light Transmitted through Earth’s Atmosphere.



This graph shows wavelengths ranging from ultraviolet (far left) to infrared. The many downward “spikes” come from absorption of particular wavelengths by molecules in Earth’s atmosphere. Some of these compounds, like water and the combination oxygen/ozone and methane, might reveal Earth as both habitable and inhabited. We will have to rely on this sort of information to seek life on exoplanets, but our spectra will be of much poorer quality than this one, in part because we will receive so little light from the planet. (credit: modification of work by NASA)

Key Concepts and Summary

The search for life beyond Earth offers several intriguing targets. Mars appears to have been more similar to Earth during its early history than it is now, with evidence for liquid water on its ancient surface and perhaps even now below ground. The accessibility of the martian surface to our spacecraft offers the exciting potential to directly examine ancient and modern samples for evidence of life. In the outer solar system, the moons Europa and Enceladus likely host vast sub-ice oceans that may directly contact the underlying rocks—a good start in providing habitable conditions—while Titan offers a fascinating laboratory for understanding the sorts of organic chemistry that might ultimately provide materials for life. And the last decade of research on exoplanets leads us to believe that there may be

billions of habitable planets in the Milky Way Galaxy. Study of these worlds offers the potential to find biomarkers indicating the presence of life.

Glossary

biomarker
: evidence of the presence of life, especially a global indication of life on a planet that could be detected remotely (such as an unusual atmospheric composition)

habitable zone
: the region around a star in which liquid water could exist on the surface of terrestrial-sized planets, hence the most probable place to look for life in a star's planetary system

The Search for Extraterrestrial Intelligence

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain why spaceships from extraterrestrial civilizations are unlikely to have visited us
- List efforts by humankind to communicate with other civilizations via messages on spacecraft
- Understand the various SETI programs scientists are undertaking

Given all the developments discussed in this chapter, it seems likely that life could have developed on many planets around other stars. Even if that life is microbial, we saw that we may soon have ways to search for chemical biosignatures. This search is of fundamental importance for understanding biology, but it does not answer the question, “Are we alone?” that we raised at the beginning of this chapter. When we ask this question, many people think of other intelligent creatures, perhaps beings that have developed technology similar to our own. If any intelligent, technical civilizations have arisen, as has happened on Earth in the most recent blink of cosmic time, how could we make contact with them?

This problem is similar to making contact with people who live in a remote part of Earth. If students in the United States want to converse with students in Australia, for example, they have two choices. Either one group gets on an airplane and travels to meet the other, or they communicate by sending a

message remotely. Given how expensive airline tickets are, most students would probably select the message route.

In the same way, if we want to get in touch with intelligent life around other stars, we can travel, or we can try to exchange messages. Because of the great distances involved, interstellar space travel would be very slow and prohibitively expensive. The fastest spacecraft the human species has built so far would take almost 80,000 years to get to the nearest star. While we could certainly design a faster craft, the more quickly we require it to travel, the greater the energy cost involved. To reach neighboring stars in less than a human life span, we would have to travel close to the speed of light. In that case, however, the expense would become truly astronomical.

Interstellar Travel

Bernard Oliver, an engineer with an abiding interest in life elsewhere, made a revealing calculation about the costs of rapid interstellar space travel. Since we do not know what sort of technology we (or other civilizations) might someday develop, Oliver considered a trip to the nearest star (and back again) in a spaceship with a “perfect engine”—one that would convert its fuel into energy with 100% efficiency. Even with a perfect engine, the energy cost of a single round-trip journey at 70% the speed of light turns out to be equivalent to several hundred thousand years’ worth of total U.S. electrical energy consumption. The cost of such travel is literally out of this world.

This is one reason astronomers are so skeptical about claims that UFOs are spaceships from extraterrestrial civilizations. Given the distance and energy expense involved, it seems unlikely that the dozens of UFOs (and even UFO abductions) claimed each year could be visitors from other stars so fascinated by Earth civilization that they are willing to expend fantastically large amounts of energy or time to reach us. Nor does it seem credible that these visitors have made this long and expensive journey and then systematically avoided contacting our governments or political and intellectual leaders.

Not every UFO report has been explained (in many cases, the observations are sketchy or contradictory). But investigation almost always converts them to IFOs (identified flying objects) or NFOs (not-at-all flying objects). While some are hoaxes, others are natural phenomena, such as bright planets, ball lightning, fireballs (bright meteors), or even flocks of birds that landed in an oil slick to make their bellies reflective. Still others are human craft, such as private planes with some lights missing, or secret military aircraft. It is also interesting that the group of people who most avidly look at the night sky, the amateur astronomers, have never reported UFO sightings. Further, not a single UFO has ever left behind any physical evidence that can be tested in a laboratory and shown to be of nonterrestrial origin.

Another common aspect of belief that aliens are visiting Earth comes from people who have difficulty accepting human accomplishments. There are many books and TV shows, for example, that assert that humans could not have built the great pyramids of Egypt, and therefore they must have been built by aliens. The huge statues (called Moai) on Easter Island are also sometimes claimed to have been built by aliens. Some people even think that the accomplishments of space exploration today are based on alien technology.

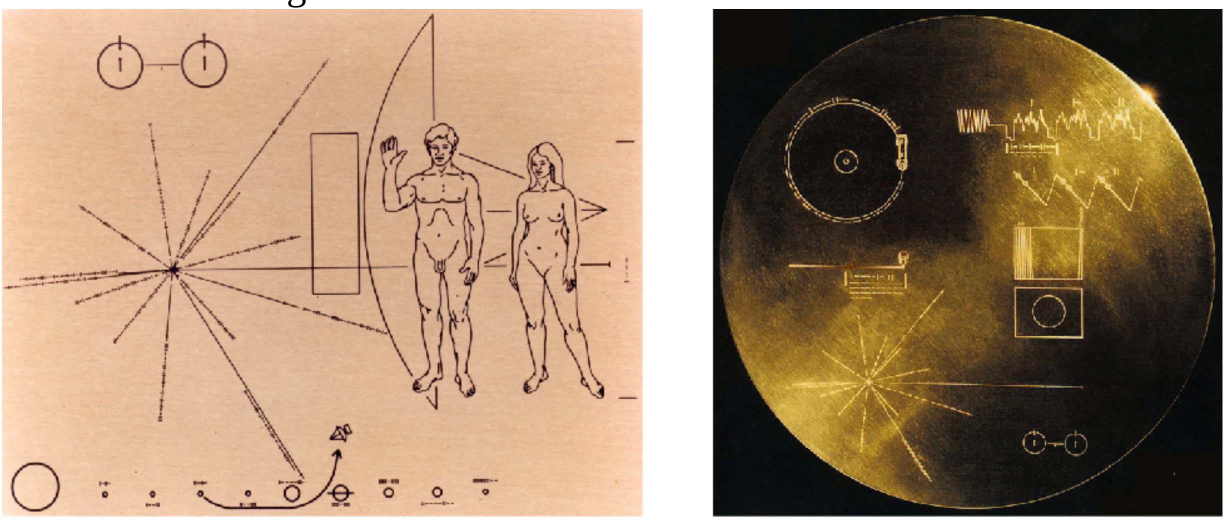
However, the evidence from archaeology and history is clear: ancient monuments were built by ancient *people*, whose brains and ingenuity were every bit as capable as ours are today, even if they didn’t have electronic textbooks like you do.

Messages on Spacecraft

While space travel by living creatures seems very difficult, robot probes can travel over long distances and over long periods of time. Five spacecraft—two Pioneers, two Voyagers, and New Horizons—are now leaving the solar system. At their coasting speeds, they will take hundreds of thousands or millions of years to get anywhere close to another star. On the other hand, they were the first products of human technology to go beyond our home system, so we wanted to put messages on board to show where they came from.

Each Pioneer carries a plaque with a pictorial message engraved on a gold-anodized aluminum plate ([link]). The Voyagers, launched in 1977, have audio and video records attached, which allowed the inclusion of over 100 photographs and a selection of music from around the world. Given the enormous space between stars in our section of the Galaxy, it is very unlikely that these messages will ever be received by anyone. They are more like a note in a bottle thrown into the sea by a shipwrecked sailor, with no realistic expectation of its being found soon but a slim hope that perhaps someday, somehow, someone will know of the sender's fate.

Interstellar Messages.



(a) (b)

(a) This is the image engraved on the plaques aboard the Pioneer 10 and 11 spacecraft. The human figures are drawn in proportion to the spacecraft, which is shown behind them. The Sun and planets in the solar system can be seen at the bottom, with the trajectory that the spacecraft followed. The lines and markings in the left center show the positions and pulse periods for a number of pulsars, which might help locate the spacecraft's origins in space and time. (b) Encoded onto a gold-coated copper disk, the Voyager record contains 118 photographs, 90 minutes of music from around the world, greetings in almost 60 languages, and other audio material. It is a summary of the sights and sounds of Earth. (credit a, b: modification of work by NASA)

Note:
The Voyager Message
An Excerpt from the Voyager Record:
"We cast this message into the cosmos. It is likely to survive a billion years into our future, when our civilization is profoundly altered. . . . If [another] civilization intercepts Voyager and can understand these recorded contents, here is our message:
This is a present from a small, distant world, a token of our sounds, our science, our images, our music, our thoughts, and our feelings. We are attempting to survive our time so we may live into yours. We hope, someday, having solved the problems we face, to join a community of galactic civilizations. This record represents our hope and our determination, and our goodwill in a vast and awesome universe."
—Jimmy Carter, President of the United States of America, June 16, 1977

Communicating with the Stars

If direct visits among stars are unlikely, we must turn to the alternative for making contact: exchanging messages. Here the news is a lot better. We already use a messenger—light or, more generally, electromagnetic waves —that moves through space at the fastest speed in the universe. Traveling at 300,000 kilometers per second, light reaches the nearest star in only 4 years and does so at a tiny fraction of the cost of sending material objects. These advantages are so clear and obvious that we assume they will occur to any other species of intelligent beings that develop technology.

However, we have access to a wide spectrum of electromagnetic radiation, ranging from the longest-wavelength radio waves to the shortest-wavelength gamma rays. Which would be the best for interstellar communication? It would not be smart to select a wavelength that is easily absorbed by interstellar gas and dust, or one that is unlikely to penetrate the atmosphere of a planet like ours. Nor would we want to pick a wavelength that has lots of competition for attention in our neighborhood.

One final criterion makes the selection easier: we want the radiation to be inexpensive enough to produce in large quantities. When we consider all these requirements, radio waves turn out to be the best answer. Being the lowest-frequency (and lowest-energy) band of the spectrum, they are not very expensive to produce, and we already use them extensively for communications on Earth. They are not significantly absorbed by interstellar dust and gas. With some exceptions, they easily pass through Earth's atmosphere and through the atmospheres of the other planets we are acquainted with.

The Cosmic Haystack

Having made the decision that radio is the most likely means of communication among intelligent civilizations, we still have many questions and a daunting task ahead of us. Shall we *send* a message, or try to *receive* one? Obviously, if every civilization decides to receive only, then no one will be sending, and everyone will be disappointed. On the other hand, it may be appropriate for us to *begin* by listening, since we are likely to be among the most primitive civilizations in the Galaxy who are interested in exchanging messages.

We do not make this statement to insult the human species (which, with certain exceptions, we are rather fond of). Instead, we base it on the fact that humans have had the ability to receive (or send) a radio message across interstellar distances for only a few decades. Compared to the ages of the stars and the Galaxy, this is a mere instant. If there are civilizations out there that are ahead of us in development by even a short time (in the cosmic sense), they are likely to have a technology head start of many, many years.

In other words, we, who have just started, may well be the "youngest" species in the Galaxy with this capability (see the discussion in [link]). Just as the youngest members of a community are often told to be quiet and listen to their elders for a while before they say something foolish, so may we want to begin our exercise in extraterrestrial communication by listening.

Even restricting our activities to listening, however, leaves us with an array of challenging questions. For example, if an extraterrestrial civilization's signal is too weak to be detected by our present-day radio telescopes, we will not detect them. In addition, it would be very expensive for an extraterrestrial civilization to broadcast on a huge number of channels. Most likely, they select one or a few channels for their particular message. Communicating on a narrow band of channels also helps distinguish an artificial message from the radio static that comes from natural cosmic processes. But the radio band contains an astronomically large number of possible channels. How can we know in advance which one they have selected, and how they have coded their message into the signal?

[link] summarizes these and other factors that scientists must grapple with when trying to tune in to radio messages from distant civilizations. Because their success depends on either guessing right about so many factors or searching through all the possibilities for each factor, some scientists have compared their quest to looking for a needle in a haystack. Thus, they like to say that the list of factors in [link] defines the *cosmic haystack problem*.

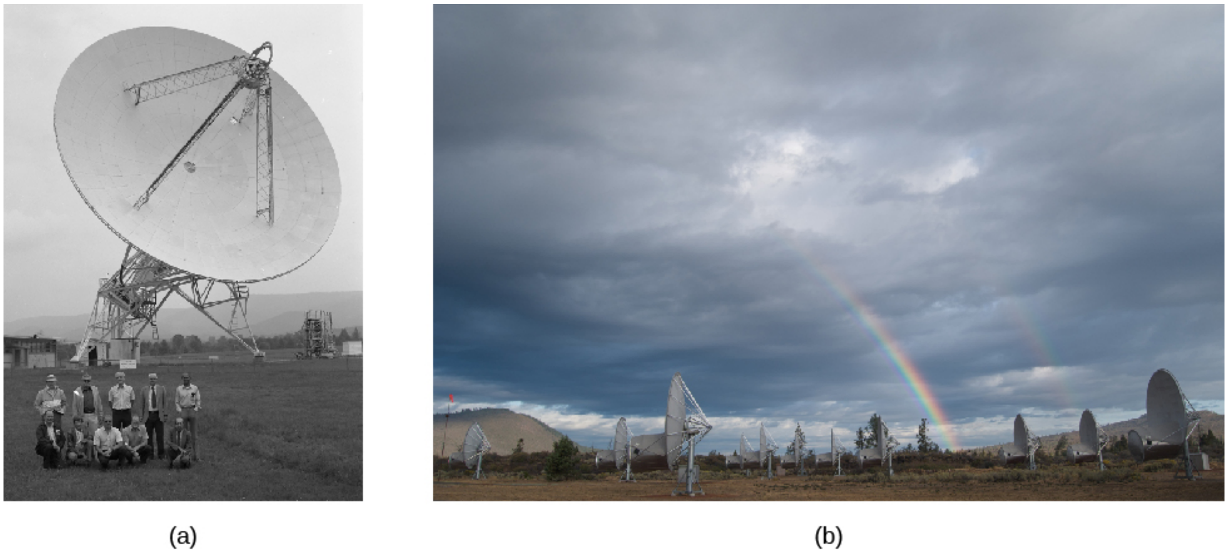
The Cosmic Haystack Problem: Some Questions about an Extraterrestrial Message
Factors
From which direction (which star) is the message coming?
On what channels (or frequencies) is the message being broadcast?
How wide in frequency is the channel?
How strong is the signal (can our radio telescopes detect it)?

The Cosmic Haystack Problem: Some Questions about an Extraterrestrial Message
Factors
Is the signal continuous, or does it shut off at times (as, for example, a lighthouse beam does when it turns away from us)?
Does the signal drift (change) in frequency because of the changing relative motion of the source and the receiver?
How is the message encoded in the signal (how do we decipher it)?
Can we even recognize a message from a completely alien species? Might it take a form we don't at all expect?

Radio Searches

Although the cosmic haystack problem seems daunting, many other research problems in astronomy also require a large investment of time, equipment, and patient effort. And, of course, if we don't search, we're sure not to find anything.

The very first search was conducted by astronomer Frank Drake in 1960, using the 85-foot antenna at the National Radio Astronomy Observatory ([link]). Called Project Ozma, after the queen of the exotic Land of Oz in the children's stories of L. Frank Baum, his experiment involved looking at about 7200 channels and two nearby stars over a period of 200 hours. Although he found nothing, Drake demonstrated that we had the technology to do such a search, and set the stage for the more sophisticated projects that followed.
Project Ozma and the Allen Telescope Array.



(a) This 25th anniversary photo shows some members of the Project Ozma team standing in front of the 85-foot radio telescope with which the 1960 search for extraterrestrial messages was performed. Frank Drake is in the back row, second from the right. (b) The Allen Telescope Array in California is made up of 42 small antennas linked together. This system allows simultaneous observations of multiple sources with millions of separate frequency channels. (credit a: modification of work by NRAO; credit b: modification of work by Colby Gutierrez-Kraybill)

Receivers are constantly improving, and the sensitivity of SETI programs—**SETI** stands for the search for extraterrestrial intelligence—is advancing rapidly. Equally important, modern electronics and software allow simultaneous searches on millions of frequencies (channels). If we can thus cover a broad frequency range, the cosmic haystack problem of guessing the right frequency largely goes away. One powerful telescope array (funded with an initial contribution from Microsoft founder Paul Allen) that is built for SETI searches is the Allen Telescope in Northern California. Other radio telescopes being used for such searches include the giant Arecibo radio dish in Puerto Rico, the recently completed, and even larger, FAST dish in China, and the Green Bank Telescope in West Virginia, which is the largest steerable radio telescope in the world.

What kind of signals do we hope to pick up? We on Earth are inadvertently sending out a flood of radio signals, dominated by military radar systems. This is a kind of leakage signal, similar to the wasted light energy that is beamed upward by poorly designed streetlights and advertising signs. Could we detect a similar leakage of radio signals from another civilization? The answer is just barely, but only for the nearest stars. For the most part, therefore, current radio SETI searches are looking for beacons, assuming that civilizations might be intentionally drawing attention to themselves or perhaps sending a message to another world or outpost that lies in our direction. Our prospects for success depend on how often civilizations arise, how long they last, and how patient they are about broadcasting their locations to the cosmos.

Note:

Jill Tarter: Trying to Make Contact

1997 was quite a year for Jill Cornell Tarter ([link]), one of the world's leading scientists in the SETI field. The SETI Institute announced that she would be the recipient of its first endowed chair (the equivalent of an endowed research professorship) named in honor of Bernard Oliver. The National Science Foundation approved a proposal by a group of scientists and educators she headed to develop an innovative hands-on high school curriculum based on the ideas of cosmic evolution (the topics of this chapter). And, at roughly the same time, she was being besieged with requests for media interviews as news reports identified her as the model for Ellie Arroway, the protagonist of *Contact*, Carl Sagan's best-selling novel about SETI. The book had been made into a high-budget science fiction film, starring Jodie Foster, who had talked with Tarter before taking the role.

Jill Tarter (credit: Christian Schidlowski)

Tarter is quick to point out, “Carl Sagan wrote a book about a woman who does what I do, not about me.” Still, as the only woman in such a senior position in the small field of SETI, she was the center of a great deal of public attention. (However, colleagues and reporters pointed out that this was nothing compared to what would happen if her search for radio signals from other civilizations recorded a success.)

Being the only woman in a group is not a new situation to Tarter, who often found herself the only woman in her advanced science or math classes. Her father had encouraged her, both in her interest in science and her “tinkering.” As an undergraduate at Cornell University, she majored in engineering physics. That training became key to putting together and maintaining the complex systems that automatically scan for signals from other civilizations.

Switching to astrophysics for her graduate studies, she wrote a PhD thesis that, among other topics, considered the formation of failed stars—those whose mass was not sufficient to ignite the nuclear reactions that power more massive stars like our own Sun. Tarter coined the term “brown

dwarf” for these small, dim objects, and it has remained the name astronomers use ever since.

It was while she was still in graduate school that Stuart Bowyer, one of her professors at the University of California, Berkeley, asked her if she wanted to be involved in a small experiment to siphon off a bit of radiation from a radio telescope as astronomers used it year in and year out and see if there was any hint of an intelligently coded radio message buried in the radio noise. Her engineering and computer programming skills became essential to the project, and soon she was hooked on the search for life elsewhere.

Thus began an illustrious career working full time searching for what she calls the technosignatures (signs of technology) from extraterrestrial civilizations. Tarter has received many awards, including being elected fellow of the American Association for the Advancement of Science in 2002, the Adler Planetarium Women in Space Science Award in 2003, and a 2009 TED Prize.

Note:

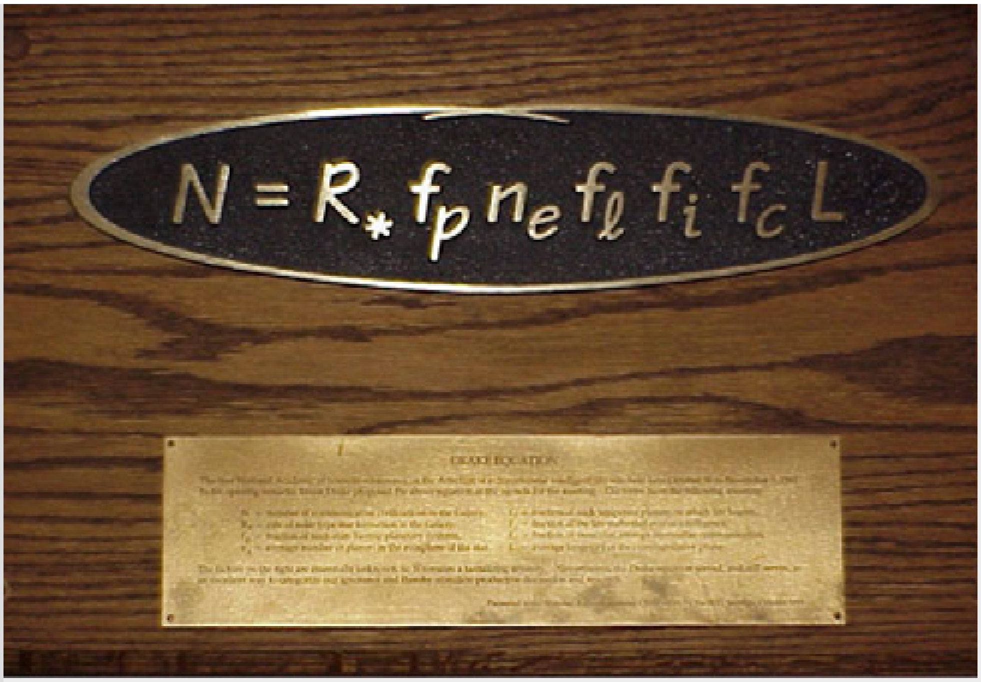
Watch the TED talk Jill Tarter gave on the fascination of the search for intelligence.

Example:

The Drake Equation

At the first scientific meeting devoted to SETI, Frank Drake wrote an equation on the blackboard that took the difficult question of estimating the number of civilizations in the Galaxy and broke it down into a series of smaller, more manageable questions. Ever since then, both astronomers and students have used this **Drake equation** as a means of approaching the most challenging question: How likely is it that we are alone? Since this is at present an unanswerable question, astronomer Jill Tarter has called the Drake equation a “way of organizing our ignorance.” (See [link].)

Drake Equation.



A plaque at the National Radio Astronomy Observatory commemorates the conference where the equation was first discussed. (credit: NRAO/NSF/AUI)

The form of the Drake equation is very simple. To estimate the number of communicating civilizations that currently exist in the Galaxy (we will define these terms more carefully in a moment), we multiply the rate of formation of such civilizations (number per year) by their average lifetime (in years). In symbols,

Equation:

$$N = R_{\text{total}} \times L$$

To make this formula easier to use (and more interesting), however, Drake separated the rate of formation R_{total} into a series of probabilities:

Equation:

$$R_{\text{total}} = R_{\text{star}} \times f_{\text{p}} \times f_{\text{e}} \times f_{\text{l}} \times f_{\text{i}} \times f_{\text{c}}$$

R_{star} is the rate of formation of stars like the Sun in our Galaxy, which is, very roughly, around 10 stars per year. Each of the other terms is a fraction

or probability (less than or equal to 1.0), and the product of all these probabilities is itself the total probability that each star will have an intelligent, technological, communicating civilization that we might want to talk to. We have:

- f_p = the fraction of these stars with planets
- f_e = the fraction of the planetary systems that include habitable planets
- f_l = the fraction of habitable planets that actually support life
- f_i = the fraction of inhabited planets that develop advanced intelligence
- f_c = the fraction of these intelligent civilizations that develop science and the technology to build radio telescopes and transmitters

Each of these factors can be discussed and perhaps evaluated, but we must guess at many of the values. In particular, we don't know how to calculate the probability of something that happened once on Earth but has not been observed elsewhere—and these include the development of life, of intelligent life, and of technological life (the last three factors in the equation). One important advance in estimating the terms of the Drake equation comes from the recent discovery of exoplanets. When the Drake equation was first written, no one had any idea whether planets and planetary systems were common. Now we know they are—another example of the Copernican principle.

Solution

Even if we don't know the answers, we can make some guesses and calculate the resulting number *N*. Let's start with the optimism implicit in the Copernican principle and set the last three terms equal to 1.0. If *R* is 10 stars/year and if we measure the average lifetime of a technological civilization in years, the units of years cancel. If we also assume that f_p is 0.1, and f_e is 1.0, the equation becomes

Equation:

$$N = R_{\text{total}} \times L = L$$

Now we see the importance of the term L, the lifetime of a communicating civilization (measured in years). We have had this capability (to communicate at the distances of the stars) for only a few decades.

Check Your Learning

Suppose we assume that this stage in our history lasts only one century.

Note:
Answer:
With our optimistic assumptions about the other factors, L = 100 years and N = 100 such civilizations in the entire Galaxy. In that case, there are so few other civilizations like ours that we are unlikely to detect any signals in a SETI search. But suppose the average lifetime is a million years; in that case, there are a million such civilizations in the Galaxy, and some of them may be within range for radio communication.
The most important conclusion from this calculation is that even if we are extremely optimistic about the probabilities, the only way we can expect success from SETI is if other civilizations are much older (and hence probably much more advanced) than ours.

Note:
Read Frank Drake's own account of how he came up with his "equation." And here is a recent interview with Frank Drake by one of the authors of this textbook.

SETI outside the Radio Realm

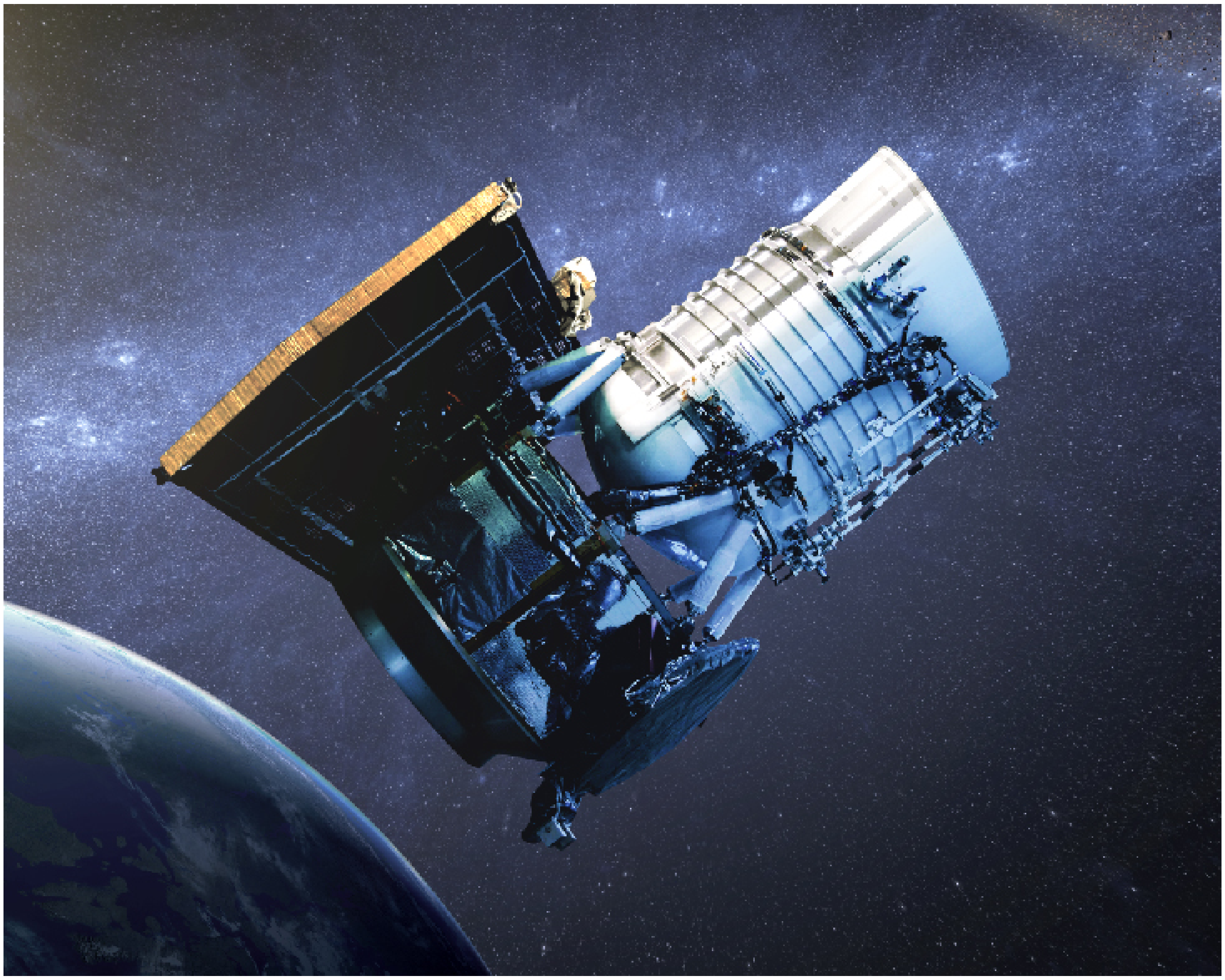
For the reasons discussed above, most SETI programs search for signals at radio wavelengths. But in science, if there are other approaches to answering an unsolved question, we don't want to neglect them. So astronomers have been thinking about other ways we could pick up evidence for the existence of technologically advanced civilizations.

Recently, technology has allowed astronomers to expand the search into the domain of visible light. You might think that it would be hopeless to try to detect a flash of visible light from a planet given the brilliance of the star it orbits. This is why we usually cannot measure the reflected light of planets around other stars. The feeble light of the planet is simply swamped by the "big light" in the neighborhood. So another civilization would need a mighty strong beacon to compete with their star.

However, in recent years, human engineers have learned how to make flashes of light brighter than the Sun. The trick is to "turn on" the light for a very brief time, so that the costs are manageable. But ultra-bright, ultra-short laser pulses (operating for periods of a billionth of a second) can pack a lot of energy and can be coded to carry a message. We also have the technology to detect such short pulses—not with human senses, but with special detectors that can be "tuned" to hunt automatically for such short bursts of light from nearby stars.

Why would any civilization try to outshine its own star in this way? It turns out that the cost of sending an ultra-short laser pulse in the direction of a few promising stars can be less than the cost of sweeping a continuous radio message across the whole sky. Or perhaps they, too, have a special fondness for light messages because one of their senses evolved using light. Several programs are now experimenting with "optical SETI" searches, which can be done with only a modest telescope. (The term *optical* here means using visible light.)

If we let our imaginations expand, we might think of other possibilities. What if a truly advanced civilization should decide to (or need to) renovate its planetary system to maximize the area for life? It could do so by breaking apart some planets or moons and building a ring of solid material that surrounds or encloses the star and intercepts some or all of its light. This huge artificial ring or sphere might glow very brightly at infrared wavelengths, as the starlight it receives is eventually converted to heat and re-radiated into space. That infrared radiation could be detected by our instruments, and searches for such infrared sources are also underway ([link]).
Wide-Field Infrared Survey Explorer (WISE).



Astronomers have used this infrared satellite to search for infrared signatures of enormous construction projects by very advanced civilizations, but their first survey did not reveal any. (credit: modification of work by NASA/JPL-Caltech)

Should We Transmit in Addition to Listening?

Our planet has some leakage of radio waves into space, from FM radio, television, military radars, and communication between Earth and our orbiting spacecraft. However, such leakage radiation is still quite weak, and therefore difficult to detect at the distances of the stars, at least with the radio technology we have. So at the present time our attempts to

communicate with other civilizations that may be out there mostly involve trying to receive messages, but not sending any ourselves.

Some scientists, however, think that it is inconsistent to search for beacons from other civilizations without announcing our presence in a similar way. (We discussed earlier the problem that if every other civilization confined itself to listening, no one would ever get in touch.) So, should we be making regular attempts at sending easily decoded messages into space? Some scientists warn that our civilization is too immature and defenseless to announce ourselves at this early point in our development. The decision whether to transmit or not turns out to be an interesting reflection of how we feel about ourselves and our place in the universe.

Discussions of transmission raise the question of who should speak for planet Earth. Today, anyone and everyone can broadcast radio signals, and many businesses, religious groups, and governments do. It would be a modest step for the same organizations to use or build large radio telescopes and begin intentional transmissions that are much stronger than the signals that leak from Earth today. And if we intercept a signal from an alien civilization, then the issue arises whether to reply.

Who should make the decision about whether, when, and how humanity announces itself to the cosmos? Is there freedom of speech when it comes to sending radio messages to other civilizations? Do all the nations of Earth have to agree before we send a signal strong enough that it has a serious chance of being received at the distances of the stars? How our species reaches a decision about these kinds of questions may well be a test of whether or not there is intelligent life on Earth.

Conclusion

Whether or not we ultimately turn out to be the only intelligent species in our part of the Galaxy, our exploration of the cosmos will surely continue. An important part of that exploration will still be the search for biomarkers from inhabited planets that have not produced technological creatures that send out radio signals. After all, creatures like butterflies and dolphins may

never build radio antennas, but we are happy to share our planet with them and would be delighted to find their counterparts on other worlds.

Whether or not life exists elsewhere is just one of the unsolved problems in astronomy that we have discussed in this book. A humble acknowledgment of how much we have left to learn about the universe is one of the fundamental hallmarks of science. This should not, however, prevent us from feeling exhilarated about how much we have already managed to discover, and feeling curious about what else we might find out in the years to come.

Our progress report on the ideas of astronomy ends here, but we hope that your interest in the universe does not. We hope you will keep up with developments in astronomy through media and online, or by going to an occasional public lecture by a local scientist. Who, after all, can even guess all the amazing things that future research projects will reveal about both the universe and our connection with it?

Key Concepts and Summary

Some astronomers are engaged in the search for extraterrestrial intelligent life (SETI). Because other planetary systems are so far away, traveling to the stars is either very slow or extremely expensive (in terms of energy required). Despite many UFO reports and tremendous media publicity, there is no evidence that any of these are related to extraterrestrial visits. Scientists have determined that the best way to communicate with any intelligent civilizations out there is by using electromagnetic waves, and radio waves seem best suited to the task. So far, they have only begun to comb the many different possible stars, frequencies, signal types, and other factors that make up what we call the cosmic haystack problem. Some astronomers are also undertaking searches for brief, bright pulses of visible light and infrared signatures of huge construction projects by advanced civilizations. If we do find a signal someday, deciding whether to answer and what to answer may be two of the greatest challenges humanity will face.

For Further Exploration

Articles

Astrobiology

Chyba, C. “The New Search for Life in the Universe.” *Astronomy* (May 2010): 34. An overview of astrobiology and the search for life out there in general, with a brief discussion of the search for intelligence.

Dorminey, B. “A New Way to Search for Life in Space.” *Astronomy* (June 2014): 44. Finding evidence of photosynthesis on other worlds.

McKay, C., & Garcia, V. “How to Search for Life on Mars.” *Scientific American* (June 2014): 44–49. Experiments future probes could perform.

Reed, N. “Why We Haven’t Found Another Earth Yet.” *Astronomy* (February 2016): 25. On the search for smaller earthlike planets in their star’s habitable zones, and where we stand.

Shapiro, R. “A Simpler Origin of Life.” *Scientific American* (June 2007): 46. New ideas about what kind of molecules formed first so life could begin.

Simpson, S. “Questioning the Oldest Signs of Life.” *Scientific American* (April 2003): 70. On the difficulty of interpreting biosignatures in rocks and the implications for the search for life on other worlds.

SETI

Chandler, D. “The New Search for Alien Intelligence.” *Astronomy* (September 2013): 28. Review of various ways of finding other civilizations out there, not just radio wave searches.

Crawford, I. “Where Are They?” *Scientific American* (July 2000): 38. On the Fermi paradox and its resolutions, and on galactic colonization models.

Folger, T. “Contact: The Day After.” *Scientific American* (January 2011): 40–45. Journalist reports on efforts to prepare for ET signals; protocols and plans for interpreting messages; and discussions of active SETI.

Kuhn, J., et al. “How to Find ET with Infrared Light.” *Astronomy* (June 2013): 30. On tracking alien civilizations by the heat they put out.

Lubick, N. “An Ear to the Stars.” *Scientific American* (November 2002): 42. Profile of SETI researcher Jill Tarter.

Nadis, S. “How Many Civilizations Lurk in the Cosmos?” *Astronomy* (April 2010): 24. New estimates for the terms in the Drake equation.

Shostak, S. “Closing in on E.T.” *Sky & Telescope* (November 2010): 22. Nice summary of current and proposed efforts to search for intelligent life out there.

Websites

Astrobiology

Astrobiology Web: http://astrobiology.com/. A news site with good information and lots of material.

Exploring Life’s Origins: http://exploringorigins.org/index.html. A website for the Exploring Origins Project, part of the multimedia exhibit of the Boston Museum of Science. Explore the origin of life on Earth with an interactive timeline, gain a deeper knowledge of the role of RNA, “build” a cell, and explore links to learn more about astrobiology and other related information.

History of Astrobiology: https://astrobiology.nasa.gov/about/history-of-astrobiology/. By Marc Kaufman, on the NASA Astrobiology site.

Life, Here and Beyond: https://astrobiology.nasa.gov/about/. By Marc Kaufman, on the NASA Astrobiology site.

SETI

Berkeley SETI Research Center: https://seti.berkeley.edu/. The University of California group has received a $100 million grant from a Russian-American billionaire to begin the Breakthrough: Listen project, a major step forward in the number of stars and number of radio channels being searched.

Fermi Paradox: http://www.seti.org/seti-institute/project/details/fermi-paradox. Could we be alone in our part of the Galaxy or, more dramatic still, could we be the only technological society in the universe? A useful discussion.

Planetary Society: http://www.planetary.org/explore/projects/seti/. This advocacy group for exploration has several pages devoted to the search for life.

SETI Institute: http://www.seti.org. A key organization in the search for life in the universe; the institute's website is full of information and videos about both astrobiology and SETI.

SETI: http://www.skyandtelescope.com/tag/seti/. *Sky & Telescope* magazine offers good articles on this topic.

Videos

Astrobiology

Copernicus Complex: Are We Special in the Cosmos?: https://www.youtube.com/watch?v=ERp0AHYRm_Q. A video of a popular-level talk by Caleb Scharf of Columbia University (1:18:54).

Life at the Edge: Life in Extreme Environments on Earth and the Search for Life in the Universe: https://www.youtube.com/watch?v=91JQmTn0SF0. A video of a 2009 nontechnical lecture by Lynn Rothschild of NASA Ames Research Center (1:31:21).

Saturn's Moon Titan: A World with Rivers, Lakes, and Possibly Even Life: https://www.youtube.com/watch?v=bbkTJeHoOKY. A video of a 2011 talk by Chris McKay of NASA Ames Research Center (1:23:33).

SETI

Allen Telescope Array: The Newest Pitchfork for Exploring the Cosmic Haystack: https://www.youtube.com/watch?v=aqsI1HZCgUM. A 2013 popular-level lecture by Jill Tarter of the SETI Institute (1:45:55).

Search for Extra-Terrestrial Intelligence: Necessarily a Long-Term Strategy: http://www.longnow.org/seminars/02004/jul/09/the-search-for-extra-terrestrial-intelligence-necessarily-a-long-term-strategy/. 2004 talk by Jill Tarter at the Long Now Foundation (1:21:13).

The Search for Extra-Terrestrial Life with New Generation Technology: https://www.youtube.com/watch?v=R1ndk08vZqM. A briefing in 2020 with Jill Tarter, Andrew Siemion, and Victoria Meadows on how scientists are currently searching for life, including intelligent life.

Search for Intelligent Life Among the Stars: New Strategies: https://www.youtube.com/watch?v=m9WxW2ktcKU. A 2010 nontechnical talk by Seth Shostak of the SETI Institute (1:29:58).

The Breakthrough: Listen Initiative—Our Boldest Effort to Answer our Oldest Question: https://www.youtube.com/watch?v=vQ2sKwwhRgI. A talk by Andrew Siemion at Harvard, with some nontechnical and some more technical sections.

Collaborative Group Activities

A. If one of the rocks from Mars examined by a future mission to the red planet does turn out to have unambiguous signs of ancient life that formed on Mars, what does your group think would be the implications of such a discovery for science and for our view of life elsewhere? Would such a discovery have any long-term effects on your own thinking?
B. Suppose we receive a message from an intelligent civilization around another star. What does your group think the implications of this discovery would be? How would your own thinking or personal philsophy be affected by such a discovery?
C. A radio message has been received from a civilization around a star 40 light-years away, which contains (in pictures) quite a bit of information about the beings that sent the message. The president of the United States has appointed your group a high-level commission to advise whether humanity should answer the message (which was not particularly directed at us, but comes from a beacon that, like a lighthouse, sweeps out a circle in space). How would you advise the president? Does your group agree on your answer or do you also have a minority view to present?
D. If there is no evidence that UFOs are extraterrestrial visitors, why does your group think that television shows, newspapers, and movies spend so much time and effort publicizing the point of view that UFOs are craft from other worlds? Make a list of reasons. Who stands to gain by exaggerating stories of unknown lights in the sky or simply fabricating stories that alien visitors are already here?
E. Does your group think scientists should simply ignore all the media publicity about UFOs or should they try to respond? If so, how should they respond? Does everyone in the group agree?
F. Suppose your group is the team planning to select the most important sights and sounds of Earth to record and put on board the next interstellar spacecraft. What pictures (or videos) and sounds would you include to represent our planet to another civilization?
G. Let’s suppose Earth civilization has decided to broadcast a message announcing our existence to other possible civilizations among the stars. Your group is part of a large task force of scientists, communications specialists, and people from the humanities charged

with deciding the form and content of our message. What would you recommend? Make a list of ideas.

H. Think of examples of contact with aliens you have seen in movies and on TV. Discuss with your group how realistic these have been, given what you have learned in this class. Was the contact in person (through traveling) or using messages? Why do you think Hollywood does so many shows and films that are not based on our scientific understanding of the universe?
I. Go through the Drake equation with your group and decide on values for each factor in the estimate. (If you disagree on what a factor should be within the group, you can have a “minority report.”) Based on the factors, how many intelligent, communicating civilizations do you estimate to be thriving in our Galaxy right now?

Review Questions

Exercise:

Problem:

What is the Copernican principle? Make a list of scientific discoveries that confirm it.

Exercise:

Problem:

Where in the solar system (and beyond) have scientists found evidence of organic molecules?

Exercise:

Problem:

Give a short history of the atoms that are now in your little finger, going back to the beginning of the universe.

Exercise:

Problem:

What is a biomarker? Give some possible examples of biomarkers we might look for beyond the solar system.

Exercise:

Problem:

Why are Mars and Europa the top targets for the study of astrobiology?

Exercise:

Problem:

Why is traveling between the stars (by creatures like us) difficult?

Exercise:

Problem:

What are the advantages to using radio waves for communication between civilizations that live around different stars? List as many as you can.

Exercise:

Problem:

What is the “cosmic haystack problem”? List as many of its components as you can think of.

Exercise:

Problem: What is a habitable zone?

Exercise:

Problem:

Why is the simultaneous detection of methane and oxygen in an atmosphere a good indication of the existence of a biosphere on that planet?

Exercise:

Problem:

What are two characteristic properties of life that distinguish it from nonliving things?

Exercise:

Problem:

What are the three requirements that scientists believe an environment needs to supply life with in order to be considered habitable?

Exercise:

Problem:

Can you name five environmental conditions that, in their extremes, microbial life has been challenged by and has learned to survive on Earth?

Thought Questions

Exercise:

Problem:

Would a human have been possible during the first generation of stars that formed right after the Big Bang? Why or why not?

Exercise:

Problem:

If we do find life on Mars, what might be some ways to check whether it formed separately from Earth life, or whether exchanges of material between the two planets meant that the two forms of life have a common origin?

Exercise:

Problem:

What kind of evidence do you think would convince astronomers that an extraterrestrial spacecraft has landed on Earth?

Exercise:

Problem:

What are some reasons that more advanced civilizations might want to send out messages to other star systems?

Exercise:

Problem:

What are some answers to the Fermi paradox? Can you think of some that are not discussed in this chapter?

Exercise:

Problem:

Why is there so little evidence of Earth’s earliest history and therefore the period when life first began on our planet?

Exercise:

Problem:

Why was the development of photosynthesis a major milestone in the evolution of life?

Exercise:

Problem: Does all life on Earth require sunshine?

Exercise:

Problem:

Why is life unlikely to be found on the surface of Mars today?

Exercise:

Problem:

In this chapter, we identify these characteristic properties of life: life extracts energy from its environment, and has a means of encoding and replicating information in order to make faithful copies of itself. Does this definition fully capture what we think of as “life”? How might our definition be biased by our terrestrial environment?

Exercise:

Problem:

Given that no sunlight can penetrate Europa’s ice shell, what would be the type of energy that could make some form of europan life possible?

Exercise:

Problem:

Why is Saturn’s moon Enceladus such an exciting place to send a mission?

Exercise:

Problem:

In addition to an atmosphere dominated by nitrogen, how else is Saturn’s moon Titan similar to Earth?

Exercise:

Problem:

How can a planet’s atmosphere affect the width of the habitable zone in its planetary system?

Exercise:

Problem:

Why are we limited to finding life on planets orbiting other stars to situations where the biosphere has created planet-scale changes?

Figuring for Yourself

Exercise:

Problem:

Suppose astronomers discover a radio message from a civilization whose planet orbits a star 35 light-years away. Their message encourages us to send a radio answer, which we decide to do. Suppose our governing bodies take 2 years to decide whether and how to answer. When our answer arrives there, their governing bodies also take two of our years to frame an answer to us. How long after we get their first message can we hope to get their reply to ours? (A question for further thinking: Once communication gets going, should we continue to wait for a reply before we send the next message?)

Exercise:

Problem:

The light a planet receives from the Sun (per square meter of planet surface) decreases with the square of the distance from the Sun. So a planet that is twice as far from the Sun as Earth receives $(1/2)^2 = 0.25$ times (25%) as much light and a planet that is three times as far from the Sun receives $(1/3)^2 = 0.11$ times (11%) as much light. How much light is received by the moons of Jupiter and Saturn (compared to Earth), worlds which orbit 5.2 and 9.5 times farther from the Sun than Earth?

Exercise:

Problem:

Think of our Milky Way Galaxy as a flat disk of diameter 100,000 light-years. Suppose we are one of 1000 civilizations, randomly distributed through the disk, interested in communicating via radio waves. How far away would the nearest such civilization be from us (on average)?

Glossary

Drake equation
: a formula for estimating the number of intelligent, technological civilizations in our Galaxy, first suggested by Frank Drake

SETI
: the search for extraterrestrial intelligence; usually applied to searches for radio signals from other civilizations

Scientific Notation

In astronomy (and other sciences), it is often necessary to deal with very large or very small numbers. In fact, when numbers become truly large in everyday life, such as the national debt in the United States, we call them astronomical. Among the ideas astronomers must routinely deal with is that the Earth is 150,000,000,000 meters from the Sun, and the mass of the hydrogen atom is 0.00000000000000000000000000167 kilograms. No one in his or her right mind would want to continue writing so many zeros!

Instead, scientists have agreed on a kind of shorthand notation, which is not only easier to write, but (as we shall see) makes multiplication and division of large and small numbers much less difficult. If you have never used this powers-of-ten notation or scientific notation, it may take a bit of time to get used to it, but you will soon find it much easier than keeping track of all those zeros.

Writing Large Numbers

In scientific notation, we generally agree to have only one number to the left of the decimal point. If a number is not in this format, it must be changed. The number 6 is already in the right format, because for integers, we understand there to be a decimal point to the right of them. So 6 is really 6., and there is indeed only one number to the left of the decimal point. But the number 965 (which is 965.) has three numbers to the left of the decimal point, and is thus ripe for conversion.

To change 965 to proper form, we must make it 9.65 and then keep track of the change we have made. (Think of the number as a weekly salary and suddenly it makes a lot of difference whether we have $965 or $9.65.) We keep track of the number of places we moved the decimal point by expressing it as a power of ten. So 965 becomes 9.65×10^2 or 9.65 multiplied by ten to the second power. The small raised 2 is called an exponent, and it tells us how many times we moved the decimal point to the left.

Note that 10^2 also designates 10 squared, or 10×10, which equals 100. And 9.65×100 is just 965, the number we started with. Another way to look at scientific notation is that we separate out the messy numbers out front, and leave the smooth units of ten for the exponent to denote. So a number like 1,372,568 becomes 1.372568 times a million (10^6) or 1.372568 times 10 multiplied by itself 6 times. We had to move the decimal point six places to the left (from its place after the 8) to get the number into the form where there is only one digit to the left of the decimal point.

The reason we call this powers-of-ten notation is that our counting system is based on increases of ten; each place in our numbering system is ten times greater than the place to the right of it. As you have probably learned, this got started because human beings have ten fingers and we started counting with them. (It is interesting to speculate that if we ever meet intelligent life-forms with only eight fingers, their counting system would probably be a powers-of-eight notation!)

So, in the example we started with, the number of meters from Earth to the Sun is 1.5×10^{11}. Elsewhere in the book, we mention that a string 1 light-year long would fit around Earth's equator 236 million or 236,000,000 times. In scientific notation, this would become 2.36×10^{8}. Now if you like expressing things in millions, as the annual reports of successful companies do, you might like to write this number as 236×10^{6}. However, the usual convention is to have only one number to the left of the decimal point.

Writing Small Numbers

Now take a number like 0.00347, which is also not in the standard (agreed-to) form for scientific notation. To put it into that format, we must make the first part of it 3.47 by moving the decimal point three places *to the right*. Note that this motion to the right is the opposite of the motion to the left that we discussed above. To keep track, we call this change negative and put a minus sign in the exponent. Thus 0.00347 becomes 3.47×10^{-3}.

In the example we gave at the beginning, the mass of the hydrogen atom would then be written as 1.67×10^{-27} kg. In this system, one is written as 10^{0}, a tenth as 10^{-1}, a hundredth as 10^{-2}, and so on. Note that any number, no matter how large or how small, can be expressed in scientific notation.

Multiplication and Division

Scientific notation is not only compact and convenient, it also simplifies arithmetic. To multiply two numbers expressed as powers of ten, you need only multiply the numbers out front and then *add* the exponents. If there are no numbers out front, as in 100 × 100,000, then you just add the exponents (in our notation, $10^{2} \times 10^{5} = 10^{7}$). When there are numbers out front, you have to multiply them, but they are much easier to deal with than numbers with many zeros in them.

Here's an example:
Equation:

$$\left(3 \times 10^5\right) \times \left(2 \times 10^9\right) = 6 \times 10^{14}$$

And here's another example:
Equation:

$$\begin{aligned} 0.04 \times 6{,}000{,}000 &= \left(4 \times 10^{-2}\right) \times \left(6 \times 10^6\right) \\ &= 24 \times 10^4 \\ &= 2.4 \times 10^5 \end{aligned}$$

Note in the second example that when we added the exponents, we treated negative exponents as we do in regular arithmetic (−2 plus 6 equals 4). Also, notice that our first result had a 24 in it, which was not in the acceptable form, having two places to the left of the decimal point, and we therefore changed it to 2.4 and changed the exponent accordingly.

To divide, you divide the numbers out front and *subtract* the exponents. Here are several examples:
Equation:

$$\begin{aligned} \frac{1{,}000{,}000}{1000} &= \frac{10^6}{10^3} = 10^{(6-3)} = 10^3 \\ \frac{9 \times 10^{12}}{2 \times 10^3} &= 4.5 \times 10^9 \\ \frac{2.8 \times 10^2}{6.2 \times 10^5} &= 0.452 \times 10^{-3} = 4.52 \times 10^{-4} \end{aligned}$$

In the last example, our first result was not in the standard form, so we had to change 0.452 into 4.52, and change the exponent accordingly.

If this is the first time that you have met scientific notation, we urge you to practice many examples using it. You might start by solving the exercises below. Like any new language, the notation looks complicated at first but gets easier as you practice it.

Exercises

1. At the end of September, 2015, the New Horizons spacecraft (which encountered Pluto for the first time in July 2015) was 4.898 billion km from Earth. Convert this number to scientific notation. How many astronomical units is this? (An astronomical unit is the distance from Earth to the Sun, or about 150 million km.)

2. During the first six years of its operation, the Hubble Space Telescope circled Earth 37,000 times, for a total of 1,280,000,000 km. Use scientific notation to find the number of km in one orbit.
3. In a large university cafeteria, a soybean-vegetable burger is offered as an alternative to regular hamburgers. If 889,875 burgers were eaten during the course of a school year, and 997 of them were veggie-burgers, what fraction and what percent of the burgers does this represent?
4. In a 2012 Kelton Research poll, 36 percent of adult Americans thought that alien beings have actually landed on Earth. The number of adults in the United States in 2012 was about 222,000,000. Use scientific notation to determine how many adults believe aliens have visited Earth.
5. In the school year 2009–2010, American colleges and universities awarded 2,354,678 degrees. Among these were 48,069 PhD degrees. What fraction of the degrees were PhDs? Express this number as a percent. (Now go and find a job for all those PhDs!)
6. A star 60 light-years away has been found to have a large planet orbiting it. Your uncle wants to know the distance to this planet in old-fashioned miles. Assume light travels 186,000 miles per second, and there are 60 seconds in a minute, 60 minutes in an hour, 24 hours in a day, and 365 days in a year. How many miles away is that star?

Answers

1. 4.898 billion is 4.898×10^9 km. One astronomical unit (AU) is 150 million km = 1.5×10^8 km. Dividing the first number by the second, we get $3.27 \times 10^{(9-8)}$ = 3.27×10^1 AU.
2. $\frac{1.28 \times 10^9 \text{ km}}{3.7 \times 10^4 \text{ orbits}} = 0.346 \times 10^{(9-4)} = 0.346 \times 10^5 = 3.46 \times 10^4 \text{ km per orbit.}$
3. $\frac{9.97 \times 10^2 \text{ veggie burgers}}{8.90 \times 10^5 \text{ total burgers}} = 1.12 \times 10^{(2-5)} = 1.12 \times 10^{(2-5)} = 1.12 \times 10^{-3}$ (or roughly about one thousandth) of the burgers were vegetarian. Percent means per hundred. So $\frac{1.12 \times 10^{-3}}{10^{-2}} = 1.12 \times 10^{(-3-(-2))} = 1.12 \times 10^{-1}$ percent (which is roughly one tenth of one percent).
4. 36% is 36 hundredths or 0.36 or 3.6×10^{-1}. Multiply that by 2.22×10^8 and you get about $7.99 \times 10^{(-1+8)} = 7.99 \times 10^7$ or almost 80 million people who believe that aliens have landed on our planet. We need more astronomy courses to educate all those people.
5. $\frac{4.81 \times 10^4}{2.35 \times 10^6} = 2.05 \times 10^{(4-6)} = 2.05 \times 10^{-2} = \text{about } 2\%$. (Note that in these examples we are rounding off some of the numbers so that we don’t have more than 2 places after the decimal point.)

6. One light-year is the distance that light travels in one year. (Usually, we use metric units and not the old British system that the United States is still using, but we are going to humor your uncle and stick with miles.) If light travels 186,000 miles every second, then it will travel 60 times that in a minute, and 60 times that in an hour, and 24 times that in a day, and 365 times that in a year. So we have $1.86 \times 10^5 \times 6.0 \times 10^1 \times 6.0 \times 10^1 \times 2.4 \times 10^1 \times 3.65 \times 10^2$. So we multiply all the numbers out front together and add all the exponents. We get $586.57 \times 10^{10} = 5.86 \times 10^{12}$ miles in a light year (which is roughly 6 trillion miles—a heck of a lot of miles). So if the star is 60 light-years away, its distance in miles is $6 \times 10^1 \times 5.86 \times 10^{12} = 35.16 \times 10^{13} = 3.516 \times 10^{14}$ miles.

Units Used in Science

In the American system of measurement (originally developed in England), the fundamental units of length, weight, and time are the foot, pound, and second, respectively. There are also larger and smaller units, which include the ton (2240 lb), the mile (5280 ft), the rod (16 1/2 ft), the yard (3 ft), the inch (1/12 ft), the ounce (1/16 lb), and so on. Such units, whose origins in decisions by British royalty have been forgotten by most people, are quite inconvenient for conversion or doing calculations.

In science, therefore, it is more usual to use the *metric system*, which has been adopted in virtually all countries except the United States. Its great advantage is that every unit increases by a factor of ten, instead of the strange factors in the American system. The fundamental units of the metric system are:

- length: 1 meter (m)
- mass: 1 kilogram (kg)
- time: 1 second (s)

A meter was originally intended to be 1 ten-millionth of the distance from the equator to the North Pole along the surface of Earth. It is about 1.1 yd. A kilogram is the mass that on Earth results in a weight of about 2.2 lb. The second is the same in metric and American units.

Length

The most commonly used quantities of length of the metric system are the following.

Conversions

Conversions
1 kilometer (km) = 1000 meters = 0.6214 mile
1 meter (m) = 0.001 km = 1.094 yards = 39.37 inches
1 centimeter (cm) = 0.01 meter = 0.3937 inch
1 millimeter (mm) = 0.001 meter = 0.1 cm
1 micrometer (μm) = 0.000001 meter = 0.0001 cm
1 nanometer (nm) = 10^{-9} meter = 10^{-7} cm

Length

To convert from the American system, here are a few helpful factors:

- 1 mile = 1.61 km
- 1 inch = 2.54 cm

Mass

Although we don't make the distinction very carefully in everyday life on Earth, strictly speaking the kilogram is a unit of mass (measuring the quantity of matter in a body, roughly how many atoms it has,) while the pound is a unit of weight (measuring how strongly Earth's gravity pulls on a body).

The most commonly used quantities of mass of the metric system are the following.

Conversions
1 metric ton = 10^6 grams = 1000 kg (and it produces a weight of 2.205×10^3 lb on Earth)
1 kg = 1000 grams (and it produces a weight of 2.2046 lb on Earth)
1 gram (g) = 0.0353 oz (and the equivalent weight is 0.002205 lb)
1 milligram (mg) = 0.001 g

Mass

A weight of 1 lb is equivalent on Earth to a mass of 0.4536 kg, while a weight of 1 oz is produced by a mass of 28.35 g.

Temperature

Three temperature scales are in general use:

- Fahrenheit (F); water freezes at 32 °F and boils at 212 °F.
- Celsius or centigrade[footnote] (C); water freezes at 0 °C and boils at 100 °C.
  Celsius is now the name used for centigrade temperature; it has a more modern standardization but differs from the old centigrade scale by less than 0.1°.
- Kelvin or absolute (K); water freezes at 273 K and boils at 373 K.

All molecular motion ceases at about −459 °F = −273 °C = 0 K, a temperature called *absolute zero*. Kelvin temperature is measured from this lowest possible temperature, and it is the temperature scale most often used in astronomy. Kelvins have the same value as centigrade or Celsius degrees, since the difference between the freezing and boiling points of water is 100 degrees in each. (Note that we just say “kelvins,” not kelvin degrees.)

On the Fahrenheit scale, the difference between the freezing and boiling points of water is 180 degrees. Thus, to convert Celsius degrees or kelvins

to Fahrenheit degrees, it is necessary to multiply by 180/100 = 9/5. To convert from Fahrenheit degrees to Celsius degrees or kelvins, it is necessary to multiply by 100/180 = 5/9.

The full conversion formulas are:

- K = °C + 273
- °C = 0.555 × (°F – 32)
- °F = (1.8 × °C) + 32

Some Useful Constants

Physical Constants	
Name	**Value**
speed of light (c)	2.9979×10^{8} m/s
gravitational constant (G)	6.674×10^{-11} $m^3/(kg\ s^2)$
Planck's constant (h)	6.626×10^{-34} J-s
mass of a hydrogen atom (M_H)	1.673×10^{-27} kg
mass of an electron (M_e)	9.109×10^{-31} kg
Rydberg constant (R_∞)	1.0974×10^{7} m^{-1}
Stefan-Boltzmann constant (σ)	5.670×10^{-8} $J/(s \cdot m^2\ deg^4)$[footnote] deg stands for degrees Celsius or kelvins
Wien's law constant ($\lambda_{max}T$)	2.898×10^{-3} m K
electron volt (energy) (eV)	1.602×10^{-19} J
energy equivalent of 1 ton TNT	4.2×10^{9} J

Astronomical Constants	
Name	**Value**
astronomical unit (AU)	1.496×10^{11} m
Light-year (ly)	9.461×10^{15} m
parsec (pc)	3.086×10^{16} m = 3.262 light-years
sidereal year (y)	3.156×10^{7} s
mass of Earth (M_{Earth})	5.974×10^{24} kg
equatorial radius of Earth (R_{Earth})	6.378×10^{6} m
obliquity of ecliptic	23° 26’
surface gravity of Earth (g)	9.807 m/s^2
escape velocity of Earth (v_{Earth})	1.119×10^{4} m/s
mass of Sun (M_{Sun})	1.989×10^{30} kg
equatorial radius of Sun (R_{Sun})	6.960×10^{8} m
luminosity of Sun (L_{Sun})	3.85×10^{26} W

Astronomical Constants	
Name	**Value**
solar constant (flux of energy received at Earth) (S)	1.368×10^3 W/m^2
Hubble constant (H_0)	approximately 20 km/s per million light-years, or approximately 70 km/s per megaparsec

The Chemical Elements

The Chemical Elements				
Element	**Symbol**	**Atomic Number**	**Atomic Weight[footnote] Where mean atomic weights have not been well determined, the atomic mass numbers of the most stable isotopes are given in parentheses.**	**Percentage of Naturally Occurring Elements in the Universe**
Hydrogen	H	1	1.008	75
Helium	He	2	4.003	23
Lithium	Li	3	6.94	6×10^{-7}
Beryllium	Be	4	9.012	1×10^{-7}
Boron	B	5	10.821	1×10^{-7}
Carbon	C	6	12.011	0.5
Nitrogen	N	7	14.007	0.1
Oxygen	O	8	15.999	1
Fluorine	F	9	18.998	4×10^{-5}
Neon	Ne	10	20.180	0.13
Sodium	Na	11	22.990	0.002

Magnesium	Mg	12	24.305	0.06
Aluminum	Al	13	26.982	0.005
Silicon	Si	14	28.085	0.07
Phosphorus	P	15	30.974	7×10^{-4}
Sulfur	S	16	32.06	0.05
Chlorine	Cl	17	35.45	1×10^{-4}
Argon	Ar	18	39.948	0.02
Potassium	K	19	39.098	3×10^{-4}
Calcium	Ca	20	40.078	0.007
Scandium	Sc	21	44.956	3×10^{-6}
Titanium	Ti	22	47.867	3×10^{-4}
Vanadium	V	23	50.942	3×10^{-4}
Chromium	Cr	24	51.996	0.0015
Manganese	Mn	25	54.938	8×10^{-4}
Iron	Fe	26	55.845	0.11
Cobalt	Co	27	58.933	3×10^{-4}
Nickel	Ni	28	58.693	0.006
Copper	Cu	29	63.546	6×10^{-6}
Zinc	Zn	30	65.38	3×10^{-5}
Gallium	Ga	31	69.723	1×10^{-6}
Germanium	Ge	32	72.630	2×10^{-5}

Arsenic	As	33	74.922	8×10^{-7}
Selenium	Se	34	78.971	3×10^{-6}
Bromine	Br	35	79.904	7×10^{-7}
Krypton	Kr	36	83.798	4×10^{-6}
Rubidium	Rb	37	85.468	1×10^{-6}
’Strontium	Sr	38	87.62	4×10^{-6}
Yttrium	Y	39	88.906	7×10^{-7}
Zirconium	Zr	40	91.224	5×10^{-6}
Niobium	Nb	41	92.906	2×10^{-7}
Molybdenum	Mo	42	95.95	5×10^{-7}
Technetium	Tc	43	(98)	—
Ruthenium	Ru	44	101.07	4×10^{-7}
Rhodium	Rh	45	102.906	6×10^{-8}
Palladium	Pd	46	106.42	2×10^{-7}
Silver	Ag	47	107.868	6×10^{-8}
Cadmium	Cd	48	112.414	2×10^{-7}
Indium	In	49	114.818	3×10^{-8}
Tin	Sn	50	118.710	4×10^{-7}
Antimony	Sb	51	121.760	4×10^{-8}
Tellurium	Te	52	127.60	9×10^{-7}
Iodine	I	53	126.904	1×10^{-7}

Xenon	Xe	54	131.293	1×10^{-6}
Cesium	Cs	55	132.905	8×10^{-8}
Barium	Ba	56	137.327	1×10^{-6}
Lanthanum	La	57	138.905	2×10^{-7}
Cerium	Ce	58	140.116	1×10^{-6}
Praseodymium	Pr	59	140.907	2×10^{-7}
Neodymium	Nd	60	144.242	1×10^{-6}
Promethium	Pm	61	(145)	—
Samarium	Sm	62	150.36	5×10^{-7}
Europium	Eu	63	151.964	5×10^{-8}
Gadolinium	Gd	64	157.25	2×10^{-7}
Terbium	Tb	65	158.925	5×10^{-8}
Dysprosium	Dy	66	162.500	2×10^{-7}
Holmium	Ho	67	164.930	5×10^{-8}
Erbium	Er	68	167.259	2×10^{-7}
Thulium	Tm	69	168.934	1×10^{-8}
Ytterbium	Yb	70	173.054	2×10^{-7}
Lutetium	Lu	71	174.967	1×10^{-8}
Hafnium	Hf	72	178.49	7×10^{-8}
Tantalum	Ta	73	180.948	8×10^{-9}
Tungsten	W	74	183.84	5×10^{-8}

Rhenium	Re	75	186.207	2×10^{-8}
Osmium	Os	76	190.23	3×10^{-7}
Iridium	Ir	77	192.217	2×10^{-7}
Platinum	Pt	78	195.084	5×10^{-7}
Gold	Au	79	196.967	6×10^{-8}
Mercury	Hg	80	200.592	1×10^{-7}
Thallium	TI	81	204.38	5×10^{-8}
Lead	Pb	82	207.2	1×10^{-6}
Bismuth	Bi	83	208.980	7×10^{-8}
Polonium	Po	84	(209)	—
Astatine	At	85	(210)	—
Radon	Rn	86	(222)	—
Francium	Fr	87	(223)	—
Radium	Ra	88	(226)	—
Actinium	Ac	89	(227)	—
Thorium	Th	90	232.038	4×10^{-8}
Protactinium	Pa	91	231.036	—
Uranium	U	92	238.029	2×10^{-8}
Neptunium	Np	93	(237)	—
Plutonium	Pu	94	(244)	—
Americium	Am	95	(243)	—

Curium	Cm	96	(247)	—
Berkelium	Bk	97	(247)	—
Californium	Cf	98	(251)	—
Einsteinium	Es	99	(252)	—
Fermium	Fm	100	(257)	—
Mendelevium	Md	101	(258)	—
Nobelium	No	102	(259)	—
Lawrencium	Lr	103	(262)	—
Rutherfordium	Rf	104	(267)	—
Dubnium	Db	105	(268)	—
Seaborgium	Sg	106	(271)	—
Bohrium	Bh	107	(272)	—
Hassium	Hs	108	(270)	—
Meitnerium	Mt	109	(276)	—
Darmstadtium	Ds	110	(281)	—
Roentgenium	Rg	111	(280)	—
Copernicium	Cn	112	(285)	—
Nihonium	Nh	113	(284)	—
Flerovium	Fl	114	(289)	—
Moskovium	Mc	115	(288)	—
Livermorium	Lv	116	(293)	—

Tennessine	Ts	117	(294)	—
Oganesson	Og	118	(294)	—

Note: Some of the newest elements near the bottom of the table have suggested names that are still under review, and those names are not yet listed here. For example, Tennessine is suggested for element 117, after the state where the Oak Ridge National Laboratory is located.